



OPEN ACCESS

EDITED BY

Roseann E. Peterson,
Suny Downstate Health Sciences University,
United States

REVIEWED BY

Jared V. Balbona,
Virginia Commonwealth University,
United States
Dovini Jayasinghe,
University of South Australia, Australia

*CORRESPONDENCE

Brad Verhulst,
✉ brad.verhulst@gmail.com

RECEIVED 03 September 2025

REVISED 06 October 2025

ACCEPTED 15 October 2025

PUBLISHED 07 November 2025

CITATION

Verhulst B (2025) Identifying gene-environment interactions across genome-wide, twin, and polygenic risk score approaches. *Front. Genet.* 16:1698381. doi: 10.3389/fgene.2025.1698381

COPYRIGHT

© 2025 Verhulst. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identifying gene-environment interactions across genome-wide, twin, and polygenic risk score approaches

Brad Verhulst*

Department of Psychiatry and Behavioral Sciences, Texas A&M University, College Station, TX, United States

Introduction: Until recently, many researchers have been hesitant to conduct genome-wide gene-environment interaction (GxE) research due to perceptions of low rates of statistical power and skepticism from controversial findings from the existing literature. Nevertheless, twin and polygenic risk score (PRS) studies suggest that GxE is pervasive and may have a large impact on complex genetic traits. Our goal in this paper is to demonstrate that consistent findings emerge from twin, PRS, and genome-wide approaches to identify GxE, subject to the known limitations for each method.

Method: We conducted a series of simulation studies, generating dataset that can be used in twin, PRS and GWAS analyses.

Results: We highlight a high degree of consistency across approaches, with each method detecting GxE. Specifically, genome-wide approaches identify individual variants that interact with an environmental moderator, but struggle with low statistical power when a trait is highly polygenic. Alternatively, aggregating genome-wide effects from a discovery sample into a PRS in the target sample increases the ability to detect broad genetic effects. However, if the statistical power in the discovery sample is low, the associations with the PRS tend to underestimate the genetic signal. This is true for both genetic main and interaction effects. Finally, twin studies are generally robust to differences in polygenicity as well as the underlying distributions of the genetic main and interaction effects. The ability of all three methods to robustly identify genomic moderation emphasizes the fact that multiple valid ways to detect GxE exist that stem from the same basic assumptions about the genetic architecture of complex traits.

KEYWORDS

gene-environment interaction (GxE), twin models, polygenic risk score (PRS), genome-wide association study (GWAS), data simulation

1 Introduction

Genetic associations often depend on contextual factors that amplify or suppress the relationships between genotypes and phenotypes. These interactions are broadly referred to as gene-environment interaction (or GxE). Genome-wide association study (GWAS) findings emphasize the incredible complexity of the relationships between genetic factors and almost every aspect of human behavior and disease. However, traditional GWAS approaches ignore GxE and the mechanisms through which contextual factors potentially regulate phenotypic outcomes (Verhulst et al., 2021). Identifying the genetic

variants that respond to contextual variation may provide critical insights into behavior modification. The majority of GxE evidence comes from twin studies, polygenic risk score (PRS) by environment interaction (PRSxE) studies, and candidate gene-environment interaction studies (which we do not discuss in detail; see Duncan and Keller, 2011). Genome-wide GxE approaches remain relatively unexplored, creating the artificial perception of methodological barriers for understanding genomic sensitivity to contextual differences. While each method has unique strengths and limitations, our overarching premise is that if GxE exists, we should be able to detect it with multiple methods, *ceteris paribus*. In this study, we present a series of simulation studies to illustrate the interconnected analytical underpinnings between twin, PRS, and genome-wide GxE models. The simulation ensures that the same data, or portions thereof, can be used for all analyses. The accompanying R script can be found on GitHub (<https://github.com/bradverhulst/GxESimulation/>), where readers can reproduce the analyses and extend them to explore nuanced variations of GxE in different scenarios.

The simulation studies we present steadily increase in complexity, from a single variant with no moderation to a highly polygenic system with extensive moderation. Despite the complexity of these simulations, real world analyses are substantially more complex. Accordingly, the simulation studies highlight potential mathematical and biological processes that could affect the statistical decision making in GxE studies, as well as problems that may arise in the interpretation of the results. The major benefit of simulation studies is that we know what the results should look like, and therefore, are able to identify how each methodological approach is affected by the data generating process.

1.1 Defining “genes” and “environments”

Biologically, a gene is a sequence of nucleotides that contain the necessary information for making a protein that contributes to the expression of one or more physical characteristics or traits (Gerstein et al., 2007). However different methods operationalize “genes” as a variety of interrelated constructs focusing on inherited biological differences. For example, in twin and family models, genes imply a latent biological predisposition for a behavior (Neale and Cardon, 1992). PRS analyses, by contrast, conceptualize genes as an aggregated or weighted sum of the relevant risk alleles for a behavior across the genome (Cross et al., 2022). Alternatively, GWASs focus on single nucleotide polymorphisms (SNPs), or small insertions or deletions (indels), and have a tendency of blurring the distinction between SNPs/indels and genes. Each of these conceptualizations of genes can interact with environmental factors to amplify or suppress the likelihood of a behavior. The implications of these distinct definitions of genes for GxE is heterogeneous, as different operationalizations of genes constrain how environments may moderate their association with a phenotype.

Defining the environment is a similarly amorphous task. In the broadest sense, an environment is anything that is not a gene. The extreme breadth of this definition may include literal environments such as the proximity to parks and greenspaces (Reed et al., 2022), inferred environments such as socioeconomic status (Abdellaoui

et al., 2025), adverse events like physical assault or maltreatment (Morrison et al., 2021), or even Petri dishes with different types of agars. More restrictive definitions require that environments are not heritable or influenced by genetic factors. However, such narrow definitions are complicated by the fact that genetic factors often cryptically contribute to variables that are typically considered environments (Kendler and Baker, 2007; McAdams et al., 2013), such as where a person lives, their social networks, and other social determinants of health and behavior. Accordingly, almost no environment satisfies the strict definition. From a GxE perspective, the fact that moderators may have a genetic component complicates the interpretation of any findings, making it difficult, if not impossible, to disaggregate pure gene-environment interactions from gene-gene interactions, or gene-moderator interactions. To circumvent this potential quagmire, when we refer to GxE, we focus on gene-moderator interactions, accepting any potential interpretational ambiguity that may be implied.

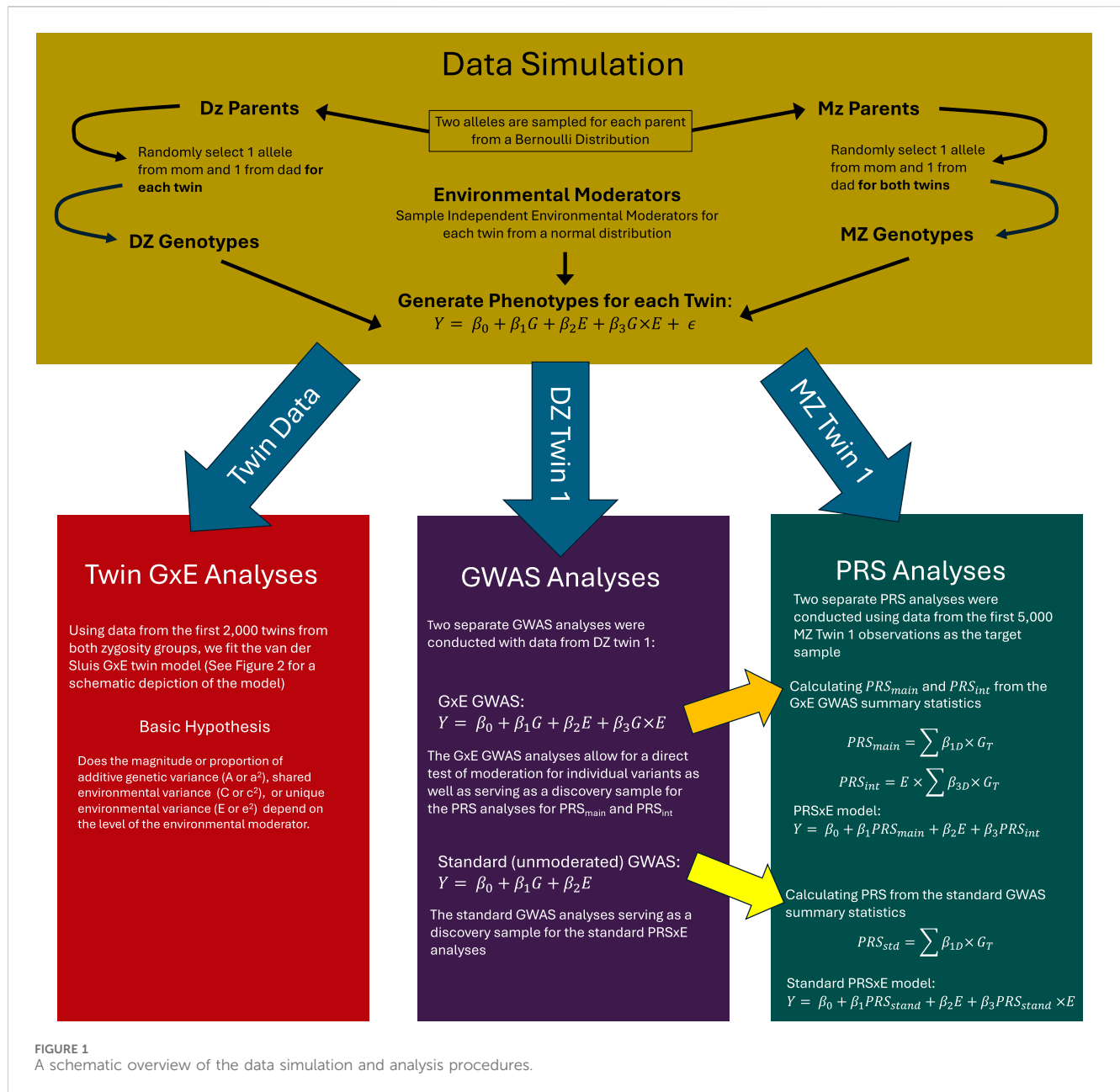
2 Methods

A series of simulation studies were conducted to examine the ability of GWAS, PRS and twin models to detect GxE. All simulations were conducted in R (v4.5.1; R Core Team, 2025) using MASS 7.3 (Venables and Ripley, 2002) for generating data. Twin analyses were conducted with OpenMx 2.22.7 (Boker et al., 2011; Neale et al., 2016). All other analyses were conducted using available functions within R. Several functions were defined to simplify the analyses, which can be found on GitHub (<https://github.com/bradverhulst/GxESimulation/>). All the code to simulate data, conduct the analyses, and plot the results are also available on GitHub. A schematic overview of the data simulation and analysis procedures is presented in Figure 1.

2.1 Data simulation procedures

Data were generated so that the same data could be simultaneously used to conduct twin, PRS and GWAS analyses. In the data simulation, we generated genotypes for 100,000 DZ mothers and fathers (and 100,000 MZ parental dyads) by randomly sampling each allele from a binomial distribution: two alleles (A1 and A2) for each parent. The maternal and paternal alleles were then summed to obtain the respective genotypes. Random normal variables were then generated to approximate the residual environmental variance for each parent. To generate the DZ genotypes, we randomly sampled one maternal and one paternal allele for each twin, which produces a correlation of $r = 0.50$ between the genotypes across the DZ twin pairs. For the MZ genotypes, we sampled one maternal and one paternal allele for both twins, which produces a correlation of $r = 1$ between the MZ genotypes. For each twin type, values for the moderators and residual variances were sampled from a normal distribution.

Continuous, quantitative phenotypes for each twin were constructed based on the algebra for the moderated regression formula (Equation 1). Regression coefficients for the direct effect of the genotypes on the phenotypes and the interaction effects were



generated under two conditions. For Studies 1-3, regression coefficients had a constant effect where the sum of the squared genetic effects was 1. In Studies 2 and 3, the sum of the squared moderated effects was also 1. For Study 4, regression coefficients were drawn from a normal distribution with a mean of zero and a variance of $(1/nSNPs)^2$. Interested users can use the data generating functions from GitHub to specify the magnitude of the direct effects of the moderator and the intercept, but this is not discussed.

2.2 Twin models

Twin GxE models were fit using structural equation modeling in OpenMx (Neale et al., 2016). In standard twin models, phenotypic variance is decomposed into additive genetic variance (V_a or a^2),

common environmental variance (V_c or c^2), and unique environmental variance (V_e or e^2). Twin GxE models extend this framework by adding a moderation parameter to each variance component that allows for the amplification or suppression of that specific source of variance: i.e., $(a + \beta_a \text{ mod})^2$, $(c + \beta_c \text{ mod})^2$, and $(e + \beta_e \text{ mod})^2$. The twin GxE models conducted here rely on the van der Sluis specification (Van Der Sluis et al., 2012). A path diagram of the van der Sluis GxE model is presented in Figure 2. In the van der Sluis GxE model, in addition to adding moderation to the A, C and E paths, both twin's phenotypes are regressed on their moderating environment as well as their co-twin's moderator to reduce the influence of gene-environment correlation on the moderation parameters. This model is more robust to parameter bias in situations where gene-environment correlation is present than comparable models. To ensure that the same data was used in

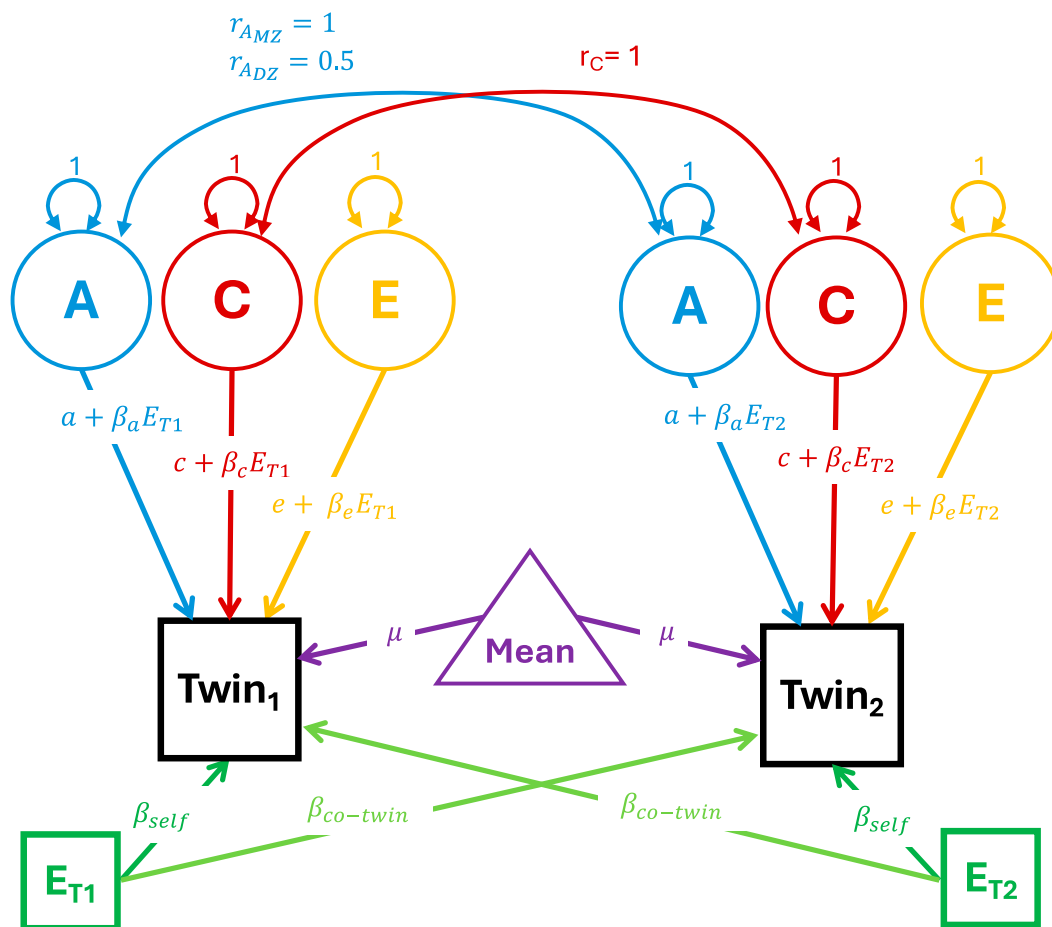


FIGURE 2

Path diagram of the Van Der Sluis et al. (2012) GxE model. Note: The phenotypic variance in the phenotypes for twins 1 and 2 is decomposed into additive genetic (A), shared environmental (C) and unique environmental E variances. Consistent with standard structural equation modeling conventions, circles indicate latent variables, squares indicate observed variables and triangles indicate means or constants. Double headed arrows indicate covariances while single headed arrows indicate regression paths. Paths with numbers (such as the variances of the latent variance components) indicate that the parameter is fixed to that value during optimization. Paths with equations or parameters indicate the estimated associations between the variables.

all analyses, we simulated data for 100,000 MZ and 100,000 DZ twin pairs. For a twin model, this sample size is beyond ridiculous. Extremely large twin registries have between 6,000 and 8,000 twin families (Lake et al., 2000; Truett et al., 1994), but other twin studies use data from about 1,000 twin families. Thus, to ensure our results are broadly consistent with the existing literature, rather than using the entire simulated sample, for all twin analyses we selected the first 2000 MZ and 2000 DZ twin pairs.

2.3 GWAS models

The GxE GWAS analyses uses the simulated data from DZ twin 1 in a standard moderated regression framework to sequentially regress the phenotype on each SNP in the genome, a moderator, and the interaction between each SNP and the moderator, as presented in Equation 1.

$$Y_i = \hat{\beta}_{0j} + \hat{\beta}_{1j}SNP_{ij} + \hat{\beta}_{2j}Env_i + \hat{\beta}_{3j}SNP_{ij}Env_i + \varepsilon_{ij} \quad (1)$$

where Y_i is a continuous phenotype and Env_i is a binary environmental moderator for the i th person, and SNP_{ij} is the j th SNP for the i th person. Accordingly, for the j th SNP, $\hat{\beta}_{0j}$ intercept, $\hat{\beta}_{1j}$ is the main or additive SNP effect, $\hat{\beta}_{2j}$ is the environmental moderator effect, and $\hat{\beta}_{3j}$ is the interaction effect. The ε_{ij} parameter captures residual variation for the j th SNP.

In addition to being of interest in their own rights, the $\hat{\beta}_{1j}$ and $\hat{\beta}_{3j}$ summary statistics from the GxE GWAS of DZ twin 1 are used as the input for the PRSxE analyses (described below).

For standard (unmoderated) GWASs, the outcome was regressed on the intercept, the environmental moderator, and each SNP in the simulated genome sequentially, as presented in Equation 2.

$$Y_i = \hat{\beta}_{0j} + \hat{\beta}_{1j}SNP_{ij} + \hat{\beta}_{2j}Env_i + \varepsilon_{ij} \quad (2)$$

where Y_i , Env_i , SNP_{ij} , and ε_{ij} are defined as above, and $\hat{\beta}_{0j}$ is the intercept, $\hat{\beta}_{1j}$ is the main or additive SNP effect and $\hat{\beta}_{2j}$ is the environmental moderator effect estimated without the interaction effect.

All GWAS significance tests were assessed using Wald tests ($Z = \beta/se$). Multiple testing corrections were based on genome-wide significance ($p < 5e-8$) and Bonferroni corrections ($\alpha = 0.05/nSNPs$). As we used simulated data, ancestry principal components and other standard covariates were not generated but should undoubtedly be included in any analyses of real data. GWASs are inherently low powered analyses, with myriad SNPs affecting the phenotype, each with a small effect size (Visscher et al., 2014). As such, the most effective and uncontroversial way of boosting power is to increase sample size. In simulation studies, this is trivially simple, limited primarily by computer memory constraints. Accordingly, we simulated GWAS data from 100,000 MZ and DZ twin families, but to simplify the statistical assumptions (specifically the independence of observations assumption), we only analyze GWAS data for DZ twin 1. These analyses become the discovery analyses for the PRSxE analyses.

2.4 PRSxE models

As PRSs require that individuals in the discovery sample are unrelated to those in the target sample, we use GWAS analyses for DZ twin 1 as the discovery dataset, and the data for the first 5,000 MZ twin 1s as the target dataset. Because statistical power in PRSs studies comes disproportionately from the sample size of the discovery sample rather than the sample size of the target sample, using 5,000 observations in the target dataset is more appropriate than the full 100,000 observation sample. Intuitively, if the regression weights are estimated more accurately in the discovery sample, the PRS values in the target sample will be more reliable and precise (Dudbridge, 2013; Palla and Dudbridge, 2015). This enhanced reliability often corresponds with larger PRS effect sizes, smaller p-values, and larger r^2 values. While these observations have been made in the context of linear PRS analyses, they should equally apply to PRSxE analyses.

Across all the simulation studies, we conducted two separate PRSxE analyses using 1) the GxE GWAS and 2) the standard (unmoderated) GWAS summary statistics as the discovery analyses. These analyses closely follow Jayasinghe et al. (2024), Werme et al. (2021), and (Choi et al., 2020). Specifically, for the PRS analyses using the GxE GWAS summary statistics as the discovery sample, we constructed PRS_{main} from the main effect parameters ($\hat{\beta}_{1Dj}$) and PRS_{int} the interaction effect parameters ($\hat{\beta}_{3Dj}$). Specifically:

$$PRS_{main} = \sum \hat{\beta}_{1Dj} SNP_{ij} \quad (3)$$

$$PRS_{int} = mod \sum \hat{\beta}_{3Dj} SNP_{ij} \quad (4)$$

Where $\hat{\beta}_{1Dj}$ and $\hat{\beta}_{3Dj}$ are the regression weights for the j th SNP in the GxE GWAS discovery analysis, and SNP_{ij} is the corresponding variant in the target sample.

The PRS analyses in the target dataset, regress the outcome (Y_i) on PRS_{main} , the environmental moderator, and PRS_{int} :

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 PRS_{main} + \hat{\beta}_2 Env_i + \hat{\beta}_3 PRS_{int} \quad (5)$$

For the PRS analyses using the unmoderated GWAS summary statistics as the discovery sample, we constructed PRS_{stand} from the main effect parameters ($\hat{\beta}_1$). Specifically:

$$PRS_{stand} = \sum \hat{\beta}_{1Dj} SNP_{ij} \quad (6)$$

Where $\hat{\beta}_{1Dj}$ is the regression weight for the j th SNP in the standard GWAS discovery analysis. The interaction effect, then, is the product between PRS_{stand} and the environmental moderator. The standard PRSxE analyses are conducted using the model:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 PRS_{stand} + \hat{\beta}_2 Env_i + \hat{\beta}_3 PRS_{stand} Env_i \quad (7)$$

2.5 Simulation study parameters

Four simulation studies were conducted to examine how different levels of polygenicity, as well as the magnitude and distribution of the effect sizes influence the ability to identify GxE. To establish baseline expectations, study 1 examines a single variant with no interaction, where a one-allele increase in a single variant increases the phenotype by 1 unit, with a residual variance of 0.50. Study 2 extends the simulation to include GxE by adding a binary (yes/no) moderator that amplifies the effect of the single variant on the phenotype by an additional 1 unit for each additional allele if the moderator is “yes”, keeping the rest of the simulated parameters the same. Study 3 further expands the simulation, generating 1,000 independent SNPs that predict the phenotype, with a uniform effect size of $\beta = \sqrt{1/1000} = 0.0316$ for each SNP. Further, we assume these SNPs interact with a binary environment at the same magnitude as the main effects. Under these conditions, the total genetic variation and moderation of the polygenic simulation mirror those of the single gene interaction analyses presented in study 2. Finally, in study 4 we conduct a more realistic simulation by simulating the distribution of the main effect and interaction coefficients from a normal distribution, allowing all variants to have a potential main effect and/or interaction effect. To maintain consistency with the previous simulations, the standard deviation of the normal distribution for the main and interaction effects was also set to $\sqrt{1/1000}$ so that the total genetic and interaction variance was approximately 1.

3 Results

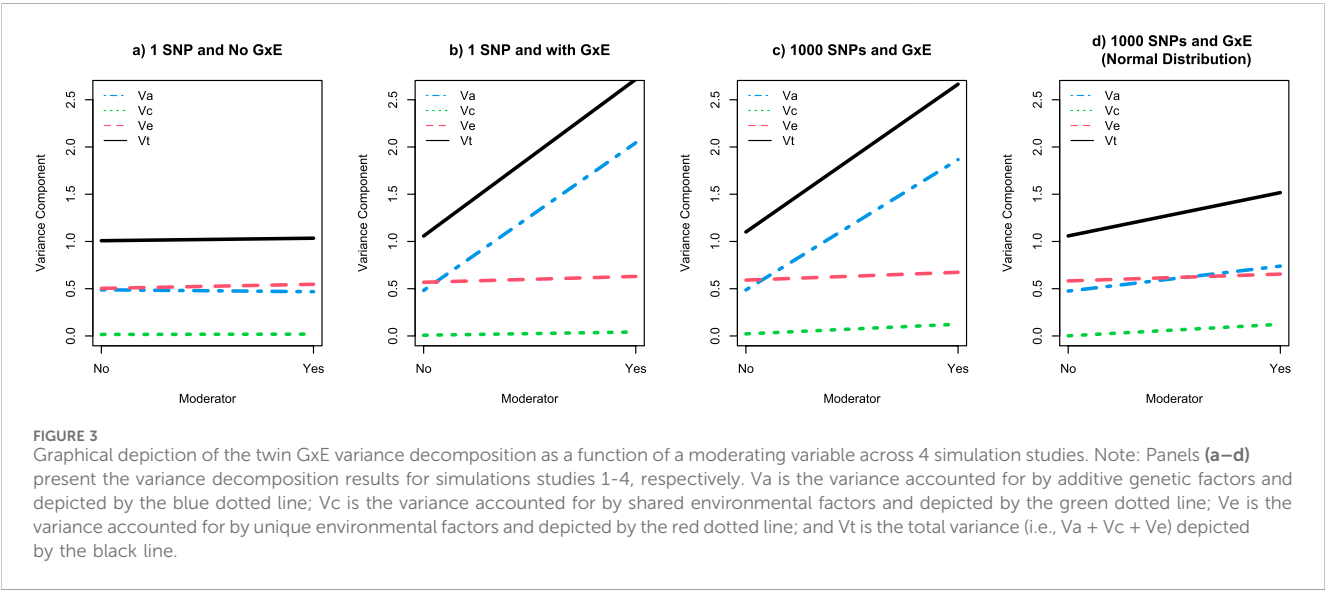
3.1 Study 1: single variant without an interaction

To begin, we generated genotypes for 100,000 MZ and DZ twin families ($N_{total} = 800,000$). As a proof of principle for the consistency across twin and genomic methods, we regressed each person's genotype on their phenotype. Consistent with the data simulation procedures, the regression coefficient for the genetic association with the phenotype is approximately 1 and the r^2 is approximately 0.5 (Table 1). While there is only one variant in this analysis, it nevertheless relies on the standard GWAS framework. If we fit a standard ACE twin model to data from 2000 MZ and 2000 DZ twins, we get estimates of $a^2 = 0.5$, $c^2 = 0$, and $e^2 = 0.5$, as shown in Figure 3a. As predicted from the biometrical model, the r^2 from the regression analyses is approximately the same as the a^2 from the twin model.

TABLE 1 Estimated regression coefficients and r^2 statistics for parents and twins with data simulated with a single variant and no interaction.

	DZ twins				MZ twins			
	Twin 1	Twin 2	Mom	Dad	Twin 1	Twin 2	Mom	Dad
Intercept	0.00	0.00	0.00	−0.01	0.00	0.00	0.00	0.00
Beta _{SNP}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
r^2	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50

The results present estimates for a simulated quantitative phenotype regressed on a simulated genetic variant for 100,000 parents and twins from MZ, and DZ, families (i.e., 800,000 individuals). Data were simulated so that a one-allele increase in a single genomic variant increased the phenotype by 1 unit, with a residual variance of 0.50 for everyone. Statistical significance measures were omitted from the table as the intercepts were all non-significant ($p > 0.05$) and the SNP, regression coefficients were all highly significant ($p < 1e-16$).



The PRS results for this data are presented in the left panels of Table 2. Because data was simulated without an interaction, in the PRSxG analyses using summary statistics from the GxE GWAS discovery models, PRS_{main} has the expected impact on the outcome and PRS_{int} has no effect. The PRSxG analyses from the standard GWAS discovery models recapitulate these findings: PRS_{stand} captures the genetic effects and the interaction effect is negligible. Furthermore, in both the GxE GWAS and standard GWAS discovery analyses, the r^2 is approximately 0.5 for both models. These results demonstrate that in the absence of GxE standard GWAS, PRS, and twin methods converge on the conclusion that genetic factors (in this case a single variant) account for ½ of the phenotypic variance.

3.2 Study 2: single variant with an interaction

We present the moderated regression analyses for a single variant and an interaction for 400,000 twins (separated into MZ and DZ twins 1 and 2). These analyses are analogous to a GxE GWAS for a single variant. As can be seen in Table 3, the regression coefficients for each twin are recovered accurately, with the main and interaction effects both having regression coefficients of approximately 1, and the intercepts and moderator effects having virtually no effect. Interestingly, the interpretation of the r^2 values for

the full model diverge from the variance accounted for in the moderator stratified analyses. Specifically, while the r^2 for the full model is approximately 0.75, when the environment is “no” the r^2 is 0.50, while the r^2 when the environment is “yes” is 0.80. Thus, the r^2 for the full model does not reflect the r^2 for either the yes or the no conditions. This is an important theme in moderation analyses, as r^2 may not be as instructive as it is in other linear regression, GWAS, or standard PRS analyses. The twin analyses, focusing on the first 2000 MZ and DZ twins, reiterate these observations. Plotting the variance components from the twin GxE model (Figure 3b) illustrates that the additive genetic variance component (Va) accounts for approximately half of the phenotypic variance (Va ≈ 0.5, Total variance or Vt ≈ 1) in the absence of the moderator, but around 80% of the phenotypic variance (Va ≈ 2, Vt ≈ 2.5) in the presence of the moderator.

We present the results of the PRSxG analyses from the GxE discovery model and the standard GWAS discovery model in the second panels of Table 2. In this simplified case, the standard GWAS discovery sample results mirror the GxE GWAS discovery analyses results, with the parameter estimates being virtually identical in both models. As with the moderated regression analyses, the main and moderation effects are approximately 1, and the intercept and the moderator effects are approximately 0. The only subtle difference between the models is that the r^2 is larger in the standard GWAS discovery sample.

TABLE 2 Results for the PRSxE analyses across the four simulation studies.

		Study 1	Study 2	Study 3	Study 4
		Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
GxE GWAS Discovery Sample	Intercept	−0.046 (0.018)*	−0.021(0.024)	−1.004 (0.598)	−0.100 (0.016)***
	PRS _{main}	1.031 (0.014)***	1.030 (0.020)**	1.048 (0.019)***	0.988 (0.014)***
	Moderator	−0.023 (0.018)	0.017 (0.035)	3.733 (0.805)***	0.077 (0.022)***
	PRS _{int}	−2.307 (2.000)	0.974 (0.028)***	0.87 (0.025)***	0.964 (0.019)***
	r ²	0.505	0.703	0.998	0.722
Standard GWAS Discovery Sample	Intercept	−0.046 (0.018)*	−0.021 (0.024)	1.387 (0.666)	−0.105 (0.021)***
	PRS _{stand}	1.031 (0.014)***	1.030 (0.020)***	0.972 (0.021)***	0.978 (0.025)***
	Moderator	−0.023 (0.018)	0.017 (0.035)	0.794 (0.944)	0.449 (0.030)***
	PRS _{stand} x moderator	0.017 (0.014)	0.976 (0.029)***	0.991 (0.030)***	−0.022 (0.035)
	r ²	0.505	0.743	0.998	0.400

Study 1 consisted of a single variant with no GxE. Study 2 had a single variant with GxE. Study 3 had 1,000 variants all with GxE. Study 4 sampled main effects and interaction effects from independent normal distributions (with a mean effect size of zero). In the GxE GWAS discovery models, PRS_{main} refers to the PRSs derived from the β_1 coefficient and PRS_{int} refers to the PRSs derived from the β_3 from the moderated regression model presented in Equations 1, 3 and 4. The regression model for PRSs derived from the GxE discovery analyses is presented in Equation 5. For the Standard GWAS discovery model, PRS_{stand} refers to the PRSs derived from β_1 from the standard GWAS presented in Equations 2 and 6, and evaluated via the regression model presented in Equation 7.

*p < .05.

**p < .01.

***p < .001.

TABLE 3 Moderated regression estimates and r² statistics for each twin in the simulation study for a single variant with an interaction.

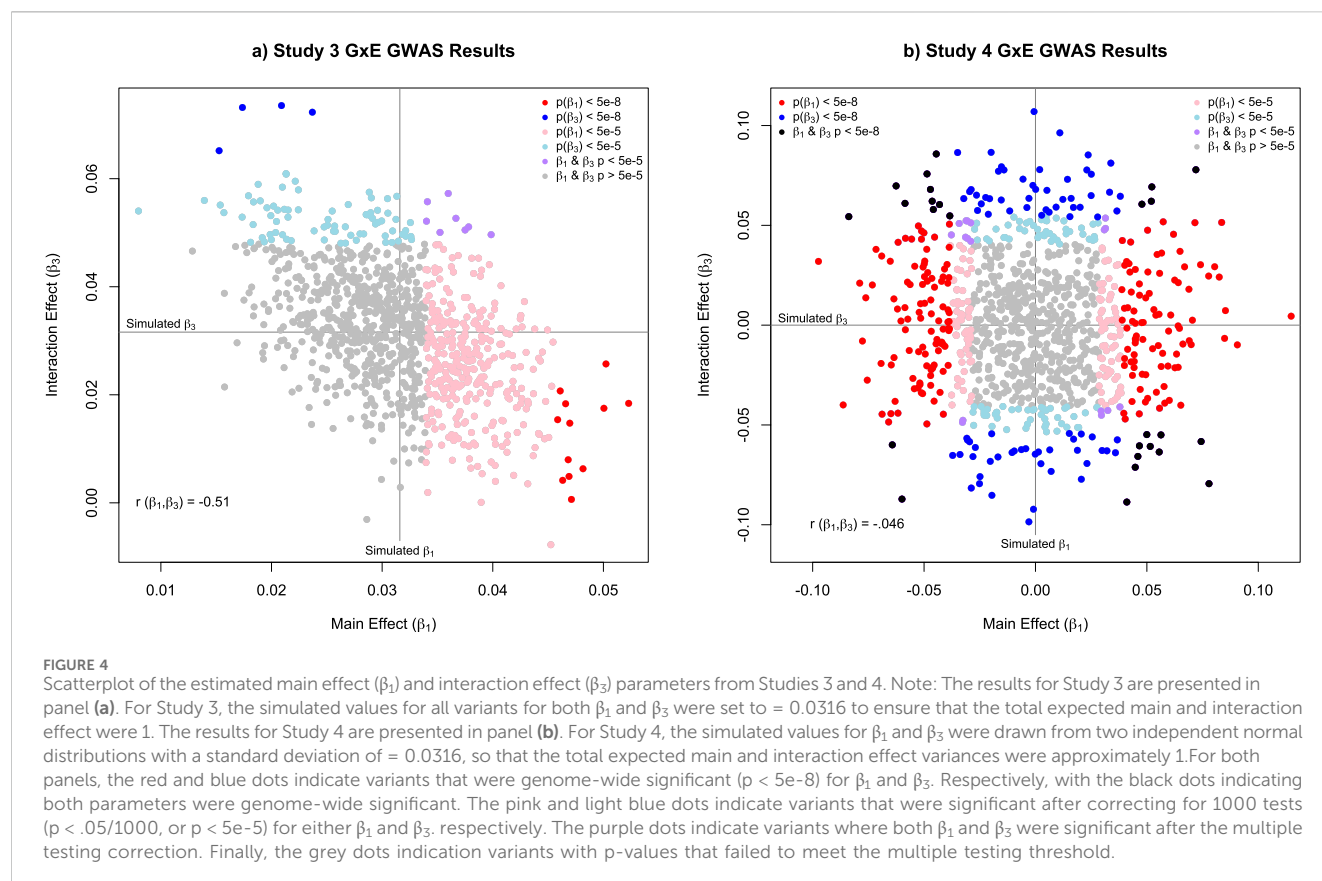
	DZ twins		MZ twins	
	Twin 1	Twin 2	Twin 1	Twin 2
Intercept	0.00	0.00	0.01	0.00
SNP	1.00	1.00	1.00	1.00
Moderator	−0.01	0.00	−0.01	0.00
SNP x Moderator	1.00	1.00	1.01	1.00
r ² (Full model)	0.75	0.75	0.75	0.75
r ² (Mod = No)	0.50	0.50	0.50	0.50
r ² (Mod = Yes)	0.80	0.80	0.80	0.80

The results present moderated regression estimates for simulated quantitative phenotypes on a simulated genetic variant for 100,000 MZ, and DZ, twins (i.e., 400,000 individuals across 4 analyses). Data were simulated so that a one-allele increase in a single genomic variant increased the phenotype by 1 unit if the moderator was 0 (“No”), and by an additional 1 unit when the moderator is 1 (“Yes”), with a residual variance of 0.50 for everyone. Statistical significance indices were omitted as the intercepts and moderator main effects were all non-significant (p > 0.05) and the SNP, regression and interaction coefficients were all highly significant (p < 1e-16).

3.3 Study 3: polygenic model with constant main effects and interactions

As complex traits are the function of myriad genetic variants, it is necessary to extend the simulation to a polygenic system with relatively small effect sizes. When examining the GxE GWAS results, it becomes immediately clear that, even with 1,000 SNPs and 100,000 observations (from the DZ twins), statistical power is an issue. Specifically, despite simulating a main effect and interaction for every SNP ($\beta_1 = \beta_3 = \sqrt{1/1000} = 0.0316$), only 12 SNPs had genome-wide significant main effects and 4 SNPs had

genome-wide significant interaction effects at the genome-wide threshold (p < 5e-8). At a more relaxed multiple testing threshold (p < 0.05/1,000) the numbers jumped to 329 significant main effects and 88 significant interaction effects. A scatterplot of the estimates for the main and interaction effects (Figure 4a) emphasizes that the significant main effects simultaneously overestimate the simulated main effect and underestimate the simulated interaction effect, and *vice versa* for the interaction effects. This is consistent with both a winner’s curse (Xiao and Boehnke, 2009) and a compensatory estimation trade-off between main and interaction effects. While it may seem strange that simulated



parameters are not significant, it reflects the reality of genetic associations.

Using only the first 2000 MZ and 2000 DZ twin pairs, the twin analysis easily identifies GxE in the polygenic GxE simulation. Interestingly, the pattern of results is almost identical to the single variant GxE simulation (Figure 3c). Like the single variant simulation, in the polygenic simulation V_a accounts for approximately half of the phenotypic variance ($V_a \approx 0.5$, $V_t \approx 1$) in the absence of the moderator and approximately 80% of the phenotypic variance ($V_a \approx 2$, $V_t \approx 2.5$) in the presence of the moderator.

Like the twin model, the PRSxE analyses for the polygenic GxE simulation are consistent with the single variant GxE simulation, with a few notable differences (third panel of Table 2). Both PRS_{main} and PRS_{int} are highly significant when the scores are generated from the GxE discovery model, but the interaction effect PRS is somewhat underestimated. When scores are derived from the standard GWAS summary statistics, we detect significant associations for both the PRS_{stand} as well as the interaction between the PRS_{stand} and the moderator.

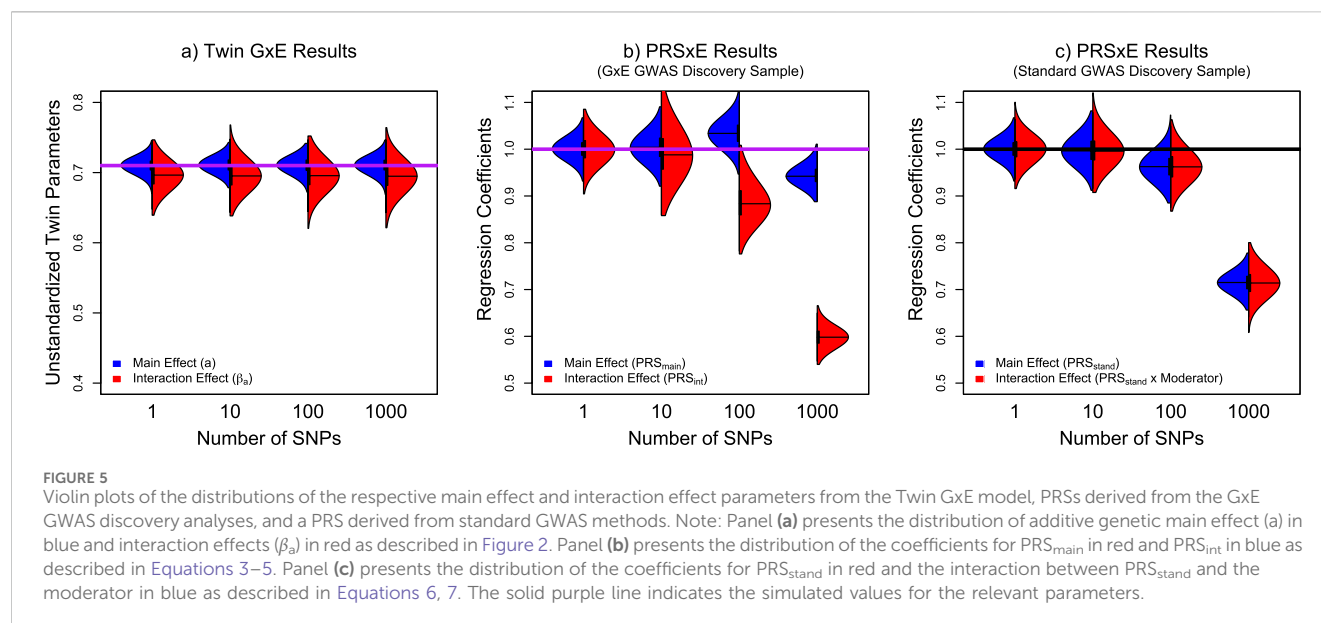
3.4 Study 4: polygenic model with random normal main effects and interactions

Up to this point, the simulations focused on models where all the variants are associated with the phenotype with the same main and interaction effects. These simulations illustrate the link between

GWAS, PRS and twin GxE models by holding the total genetic and interaction variance constant. They also lay the foundation for more complex models of polygenic moderation. In this simulation, the effect sizes for the main and interaction effects are drawn from a normal distribution with a mean of zero and a sd of $\sqrt{1/1000}$. Accordingly, while many main and/or interaction effects are effectively null, some are positive, others negative with possible effect sizes substantially larger than in study 3.

Starting with the GxE GWAS analyses for 100,000 DZ twin 1 observations, we now see 218 main effects and 106 interaction effects that are genome-wide significant, and an additional 158 main effects and 132 interaction effects that survive a multiple testing correction of $\alpha < 0.05/1000$. While these values may be larger than common effect sizes when compared to some complex polygenic traits, the results highlight the fact that when the GxE GWAS effect sizes are large, they can be easily identified (Figure 4b). As is true for moderated regression more broadly, main effects are more likely to be detected than interaction effects. This simple power discrepancy contributes to the underestimation of the prevalence of GxE and the importance of GxE for complex traits.

As with the previous simulations, the twin analysis of 2000 MZ and 2000 DZ twin families easily detects significant GxE (Figure 3d). Further, because the distributions of the main and interaction effects are independent, moderation can amplify or suppress the main effects. Nevertheless, the increase in the total variance in the simulated phenotype in the presence of the moderator is almost entirely due to the increase in genetic variation.



In study 4, the PRSxE analyses using the GxE GWAS vs. the standard GWAS as the discovery samples diverge notably (right panels of Table 2). Specifically, in analyses based on the GxE discovery sample both PRS_{main} and PRS_{int} predict the outcome at the simulated levels. For the PRSxE model using the standard GWAS as the discovery sample, PRS_{stand} is highly predictive of the outcome, but the interaction between PRS_{stand} and the moderator is trivial and not statistically significant. As there are a reasonably large number of variants with genome-wide significant GxE, the lack of an interaction signal when using the standard GWAS model as the discovery sample is unlikely. Instead, the approximately equal mixture of positive and negative interaction effects cancel out when standard PRS methods are used in the PRSxE model.

3.5 Ad hoc exploration of the underestimation of the PRS effects

In developing the simulations studies, particularly when smaller sample sizes were used, an interesting pattern emerged where we observed a consistent underestimation of the PRS effect sizes as polygenicity increased. This effect was particularly pronounced for the analyses based on the GxE GWAS discovery analyses, as can be seen in Study 3. Specifically, the PRS effect sizes in all the simulations were generated to be approximately 1. However, when relatively small sample sizes were used, the estimated PRS effects were substantially smaller than 1. As statistical power in PRS analyses depends on the precision of the discovery sample estimates (Dudbridge, 2013; Palla and Dudbridge, 2015), both the magnitude of the effects and the size of the discovery sample play a pivotal role in generating accurate PRS in the target sample. Accordingly, it is likely that the underestimated PRS effects stem from the lack of precision in the estimation of the respective coefficients in the discovery analyses, with the interaction parameters being less precise than main effects. Interestingly, the downward biased PRS effect resulted in inflated intercept and the moderator coefficients (that deviate from the simulated values of 0).

Essentially, the overestimation of these effects allowed the models to compensate for the underestimation of the PRSs.

To investigate the cause of the underestimated PRS effects, we conducted a follow-up analysis where we varied the number of simulated SNPs ($N_{snps} = 1, 10, 100, 1,000$), repeating the simulation 1,000 times for each condition with a sample size at 20,000. Thus, the 1 and 10 SNP conditions will be overpowered, but the 100 and 1,000 SNP conditions will be underpowered. As the analyses become increasingly underpowered, we should observe an increasing downward bias in the PRS estimates. As the twin parameters are sufficiently powered across all these conditions, the twin model serves as a positive control. Specifically, the additive genetic main and interaction parameters in the twin model should be approximately 0.71, and the PRSxE main and interaction parameters for both discovery samples should be 1. As shown in the left panel of Figure 5, the twin model estimates are close to the simulated parameters, with a very slight but consistent underestimation of the interaction parameter due to the inclusion of common and unique environmental moderation (see Verhulst et al., 2019 for a more complete explanation of this bias). In both PRSxE models, as the polygenicity increased, so did the underestimation. However, in models using the GxE GWAS as the discovery sample, PRS_{int} is substantially more affected than PRS_{main} . Alternatively, when using the standard GWAS as the discovery sample, because there is only PRS_{stand} , the main effect and interaction effect underestimation is symmetrical. These results suggest that PRSs derived from underpowered discovery samples will underestimate the polygenic effect.

3.6 Statistical power comparisons for GWAS main and interaction effects

Since statistical power is a constant concern in GWAS, and particularly GxE GWAS, Monte Carlo power calculations were conducted to identify the approximate sample size necessary to achieve 80% power to detect genetic associations at the genome-

TABLE 4 Power analysis results comparing the sample size and statistical power to detect main effects and interaction effects in moderated regression analyses.

Number of SNPs	Effect size	Sample size for 80% power for the main effect	Power to detect the interaction effect from the main effect power analysis	Sample size for 80% power for the interaction effect	Power to detect the main effect based on the interaction effect sample size
1	1.0	84	0.23	159	1.00
10	0.316	2,340	0.14	4,950	1.00
100	0.10	24,744	0.13	52,529	1.00
1,000	0.0316	251,674	0.13	528,683	1.00

The effect size was determined as $\sqrt{1/N_{SNPs}}$ to remain consistent with the simulation studies. The sample size for 80% power for the main and interaction effects was identified via Monte Carlo simulations and should be considered approximate. The power to detect the interaction effect (or the main effect) was based on the sample size required to achieve 80% power from the main effect (interaction effect) analysis.

wide significance level for four main and interaction effect sizes ($\beta_1 = \beta_3 = \{1.0, 0.316, 0.1, 0.0316\}$; Table 4). The power analyses highlight the extreme discrepancy in power between the β_1 and β_3 coefficients. Specifically, when the values of β_1 and β_3 are equal, it takes more than twice as many observations to achieve 80% power for β_3 as compared with β_1 , even though both parameters are estimated in the same model. Put another way, if the interaction effect size is the same as the main effect size, if an analysis has 80% power to detect the main effect, it likely has around 15% power to detect the interaction effect. Alternatively, if an analysis has 80% power to detect the interaction effect, it has almost perfect power to detect the main effect. Importantly, while β values greater than 0.1 are unrealistic for most complex traits, effect sizes in the 0.05 range are reasonable. Thus, it may take approximately 500,000 observations to achieve 80% power to detect significant β_3 coefficients for many complex traits. Importantly, when the sample size in the discovery sample has 80% power to detect the association, the underestimation of both the main and interaction effects are ameliorated.

4 Discussion

The simulation studies demonstrate substantial consistency in the ability to detect GxE across analytical approaches. The twin analyses identified increasing phenotypic heritability at riskier levels of the moderating environment. The PRSxE analyses identified both genetic main and interaction effects. The GxE GWASs identified interactions with individual genetic variants. Unsurprisingly, as we increased the polygenicity of the main and interaction effects (thereby shrinking their effect sizes), the power to detect genome-wide associations and interactions for individual variants declines. Nevertheless, both twin and PRS analyses were able to identify robust main and moderated effects with reasonably small sample sizes (i.e., 2,000 MZ and DZ twin pairs and 5,000 individuals in the PRS target sample). Therefore, twin and PRSxE analyses can be used to supplement investigations by focusing attention on (or eliminating) phenotype-moderator combinations that likely exhibit GxE.

Interestingly, while twin methods have struggled to remain relevant in the genomic era, standard twin GxE methods were extremely robust to variation in polygenicity and the direction of the interactions at the individual variant levels. For simulation studies 2 and 3, the pattern of GxE in the twin models was the

same regardless of the number of SNPs included in the data generating process. Further, the twin model effectively identified GxE even when the PRSxE models underestimated the effects or failed to identify moderation. Accordingly, twin GxE methods remain robust to many of the limitations that affect genome-wide approaches.

Across the simulation studies, it is possible to broadly conclude that the PRSxE models accurately recover the simulated values for both the main and interaction effects. However, there are three notable caveats to this conclusion. First, the pattern of individual variant GxE has a substantial impact on the PRSxE moderation effects. If the sign of the moderation was consistent across variants (e.g., the interaction effects consistently amplified the genetic associations), the PRS derived from standard GWAS summary statistics performed slightly better than PRSs derived from GxE GWAS summary statistics. However, if some interaction coefficients were positive and others negative, constructing the PRSs from standard GWAS regression weights led to the conclusion that GxE did not exist. When using PRSs derived from the GxE GWAS discovery analysis, we identified significant moderation. Effectively, using standard GWAS summary statistics to construct PRSs in a PRSxE models assumes that every SNP interacts with the environment proportional to the magnitude and direction of the discovery GWAS. If GxE amplifies the associations for some loci and suppresses others, aggregating genetic effects using standard PRS methods will effectively cancel out any moderation. Because standard GWASs do not include moderation, it is possible that using standard GWAS summary statistics to generate PRSs may lead to the conclusion that GxE does not exist, when, in fact, it does.

Second, as we demonstrate in the *post hoc* analyses, the sample size of the discovery analysis plays a critical role for identifying polygenic signatures. While this has been pointed out for linear GWASs (Dudbridge, 2013; Palla and Dudbridge, 2015), it is worth repeating in the context of GxE. If the discovery analyses are not sufficiently powered to detect the genetic associations, the regression weights will contain substantial levels of stochastic noise, and the PRSxE analyses will underestimate the genetic signal for both main and interaction effects. As interaction effects require larger samples to achieve the same level of statistical power, the underestimation of the interaction PRSs was more pronounced. Across the simulation studies, the genetic variance was held constant to enhance the comparability of the results. In real analyses every effect size is different. Therefore, in real analyses PRSs will necessarily be

TABLE 5 Summary of the hypotheses, strengths and limitations of GxE analyses using twin, PRS and GWAS methods.

GxE method	Sample requirements	GxE Hypothesis	Strengths	Limitations
Twin	MZ and DZ Twins	Does the proportion of genetic variance differ based on the level of the moderator?	• Robust to most differences in the genetic architecture of the phenotype and the pattern of GxE for individual variants • Well-validated and accepted methods with known strengths and limitations	• Twins are a relatively rare populations and require have unique recruitment barriers • Latent variables have for low levels of biological specificity making translation difficult
PRS	Discovery Sample and Target Sample	Does the PRS have different effects based on the level of the moderator?	• Data can be contributed to future GWASs and GxE GWAS meta-analyses • It is possible to stratify individual risk in ways that could be used in clinical settings (for some phenotypes) • If an appropriate discovery sample is identified, PRS target samples can be relatively small	• Requires two samples • Phenotypic and moderator heterogeneity across samples may not be assumed • Sensitive to the pattern of GxE for individual variants • GxE GWASs are rare and may not be available for the phenotype-moderator combinations of interest
GWAS	Single Sample	Do individual variants have different effects based on the level of the moderator?	• Very high levels of biological specificity • Direct extensions drug development and treatment	• Sample size or statistical power

constructed from discovery GWASs that have a mixture of effect sizes, with some associations being sufficiently powered and others underpowered. Accordingly, to the extent to which underpowered variants are included in a PRS, the results likely underestimate the importance of the genetic factors for the phenotype.

Finally, the accuracy of the r^2 statistic, which indexes the proportion of variation accounted for by the predictors in the full model, does not reflect the r^2 for specific levels of the moderator. In the single variant GxE model, the model r^2 was substantially overestimated when the moderator was absent and underestimated when it was present. In polygenic models, aggregating r^2 across a mixture of over- and underestimation in real samples may leave PRSs disconnected from any specific level of the moderator. Thus, while r^2 may be an effective metric for standard PRS analyses, its usefulness for PRSxE analyses may be limited.

Importantly, with the proliferation of high-quality, publicly accessible GWAS summary statistics, PRS and PRSxE methods are entering a phase of rapid methodological development (Ni et al., 2021). For example, methods such as PRS-CSx (Ruan et al., 2022), MegaPRS (Zhang et al., 2021), LDPred2 (Privé et al., 2021), and SBayesR (Lloyd-Jones et al., 2019) extend PRS analyses to diverse ancestry groups, re-weight the finding by including functional genomic information, and utilize powerful Bayesian methods. While we do not use these advanced PRS methods in our simulation studies, our results nevertheless can inform analyses that do. Furthermore, while our simulations do not include covariates, Type 1 Error rates can be better controlled by including higher order polynomials of the covariates (Jayasinghe et al., 2024). Thus, our findings add to the broader perspective on the accuracy of PRSs for complex traits.

The GxE GWAS simulations generally confirmed the existing knowledge regarding the low levels of statistical power and the corresponding need to have very large sample sizes to identify GxE for individual variants. However, connecting GWAS data on GxE for individual variants to twin and PRS data challenges the perception that the underlying genomic mechanisms for twin and molecular methods are categorically distinct. While twin,

PRS, and GWAS methods use different data types, rely on different statistical assumptions, and conceptualize “genes” in different ways, they can all be reduced to the same basic biological principals. Individual variant analyses, like GxE GWASs, promise extremely precise levels of biological specificity, which informs the etiology of complex traits and can be leveraged to advance personalized, genomic medicine. It will take an enormous amount of time, money, and effort to collect hundreds of thousands (and likely millions) of high-quality genotypes and phenotypes, not to mention numerous potential moderators for each phenotype. Accordingly, it is essential to use data and methods, that only require a few thousand observations, to prioritize the phenotype-moderator combinations with the highest chances of illuminating GxE for individual variants.

One of the strengths of our simulation studies is the flexibility to generate data that simultaneously satisfies the requirements of twin, GWAS and PRS analyses. However, such data is only realistically obtained within a simulation setting, as each method has substantially different requirements. Moreover, despite being able to detect GxE with each method, the research questions that can be answered with each analytical approach vary substantially. We discuss the key considerations for each method below and summarize them in Table 5.

Twin studies focus on how measured environmental moderators amplify or suppress the additive genetic, common and unique environmental variance components. Focusing on the additive genetic variance component, twin GxE models treat “genes” as a latent biological predisposition for a behavior that can be regulated by the environment. The latent biological factor effectively aggregates genome-wide genetic associations into a single variable that does not require DNA collection, instead leveraging the differences between the variances and covariances of MZ and DZ twins (Neale and Cardon, 1992; Purcell, 2002). Because latent variables cannot be measured directly and thus must be inferred from the existence of other measurable variables, twin methods cannot be used to identify actionable biological interventions. Nevertheless, twin models require substantially smaller sample

sizes to detect GxE, making them extremely valuable for identifying phenotype-moderator combinations with the potential for clinical relevance.

PRS analyses, and PRSxE studies by extension, aggregate the effect of relevant alleles for a phenotype across the genome (Cross et al., 2022). PRS analyses can effectively test whether genetic factors play a role in a behavior by projecting estimated genetic associations from the GWAS of a large discovery sample into a smaller genetically assayed target sample. PRSxE analyses refine the potential research question by asking whether environmental factors increase or decrease the importance of the genetic signal. As with twin studies, PRS and PRSxE methods cannot directly identify actionable genetic interventions, as they aggregate signals from across the genome. However, PRSs can be used to stratify a person's genetic risk of developing a disease or other phenotype. By extension, it is possible that PRSxE methods can be used to further personalize the prediction by integrating the genetic signal with an environmental moderator. If conducted appropriately, this integration can be instrumental for enhancing personalized medicine (Verhulst and Benstock, 2023). As most current PRS and PRSxE analyses utilize published GWAS summary statistics as the discovery sample, collecting or identifying data for a target sample is of critical importance. Importantly, sample sizes for these target samples can be substantially smaller than the discovery GWAS. While we used 5,000 observations in our target datasets, with sufficiently powered discovery GWAS it is possible to detect genetic signals with sample sizes in the hundreds (Chang et al., 2019).

Finally, GxE GWAS analyses provide the most biological specificity, by identifying specific SNPs that interact with an environmental moderator. However, the statistical power to detect moderated associations for individual SNPs is low. Accordingly, GxE GWASs with currently available data (including those from biobanks) may only be able to identify GxE with large moderation effects. One method of addressing power concerns in GWAS has been to develop consortia that share data and results, thereby increasing sample sizes (Adams et al., 2025; International Schizophrenia Consortium et al., 2009). Within these consortia, phenotypic heterogeneity has been a major concern for GWAS meta-analyses. Differential assessment and ascertainment strategies can obscure the interpretation of any potential associations (Cai et al., 2025). In GxE meta-analyses, not only is phenotypic heterogeneity a concern, but moderator heterogeneity will introduce additional challenges. For example, many conceptualizations of adverse life events aggregate stressors from numerous categories. While aggregated scales can increase statistical power in many situations, if individual stressors have divergent moderating effects, such effects may inhibit the detection of GxE.

Many of the lessons learned from GWASs extend to genome-wide GxE analyses, but due to the increased complexity of moderated regression over linear regression, slight extensions are required. For example, one of the central lessons from GWASs is the extreme polygenicity of complex traits: hundreds or thousands of independent loci are associated with each behavioral outcome, with individual associations exerting very small effects and thus requiring extremely large sample sizes to detect. Moderated genetic associations will likely be equally polygenic, but as interactions

are inherently less powerful, they will require even larger samples. Substantial reductions in genotyping costs over the past 2 decades have allowed for multiple historic efforts aimed at recruiting hundreds of thousands of participants assessed on a wide range of outcomes and moderators (Sudlow et al., 2015; The All of Us Research Program Investigators, 2019). Thus, the justification for eschewing genome-wide GxE due to power concerns is decreasing.

5 Conclusion

While GxE plays a critical role in the etiology of many complex traits, as demonstrated by the plethora of GxE research in twins, genome-wide GxE research has fallen behind other genomic investigations. In part, the hesitation to conduct genome-wide GxE research is a function of perceptions of low rates of statistical power and skepticism from controversial findings from the candidate gene-environment interaction literature (Culverhouse et al., 2018; Duncan and Keller, 2011). While reasonable, such perceptions have delayed genome-wide GxE investigations. The methodological consistency demonstrated by the current simulations highlight the importance of leveraging multiple methods for identifying GxE. Thus, the robust twin GxE literature should give researchers a sense of hope that corresponding findings can be identified in genome-wide data.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://github.com/bradverhulst/GxESimulation/>. Further inquiries can be directed to the corresponding author.

Author contributions

BV: Data curation, Methodology, Visualization, Investigation, Conceptualization, Project administration, Funding acquisition, Software, Formal Analysis, Resources, Writing – review and editing, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by a NARSAD grant from the Brain and Behavior Research Foundation (Award Number 31397).

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

The handling editor RP declared a past co-authorship with the author BV.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abdellouai, A., Martin, H. C., Kolk, M., Rutherford, A., Muthukrishna, M., Tropf, F. C., et al. (2025). Socio-economic status is a social construct with heritable components and genetic consequences. *Nat. Hum. Behav.* 9 (5), 864–876. doi:10.1038/s41562-025-02150-4
- Adams, M. J., Streit, F., Meng, X., Awasthi, S., Adey, B. N., Choi, K. W., et al. (2025). Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell* 188 (3), 640–652.e9. doi:10.1016/j.cell.2024.12.002
- Boker, S. M., Neale, M. C., Maes, H., Wilde, M., Spiegel, M., Brick, T. R., et al. (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika* 76 (2), 306–317. doi:10.1007/s11336-010-9200-6
- Cai, N., Verhulst, B., Andreassen, O. A., Buitelaar, J., Edenberg, H. J., Hottelma, J. M., et al. (2025). Assessment and ascertainment in psychiatric molecular genetics: challenges and opportunities for cross-disorder research. *Mol. Psychiatry* 30 (4), 1627–1638. doi:10.1038/s41380-024-02878-x
- Chang, L.-H., Couvy-Duchesne, B., Liu, M., Medland, S. E., Verhulst, B., Benetsch, E. G., et al. (2019). Association between polygenic risk for tobacco or alcohol consumption and liability to licit and illicit substance use in young Australian adults. *Drug Alcohol Dependence* 197, 271–279. doi:10.1016/j.drugalcdep.2019.01.015
- Choi, S. W., Mak, T. S.-H., and O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* 15 (9), 2759–2772. doi:10.1038/s41596-020-0353-1
- Cross, B., Turner, R., and Pirmohamed, M. (2022). Polygenic risk scores: an overview from bench to bedside for personalised medicine. *Front. Genet.* 13, 1000667. doi:10.3389/fgene.2022.1000667
- Culverhouse, R. C., Saccone, N. L., Horton, A. C., Ma, Y., Anstey, K. J., Banaschewski, T., et al. (2018). Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol. Psychiatry* 23 (1), 133–142. doi:10.1038/mp.2017.44
- Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genet.* 9 (3), e1003348. doi:10.1371/journal.pgen.1003348
- Duncan, L. E., and Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry* 168 (10), 1041–1049. doi:10.1176/appi.ajp.2011.11020191
- Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., et al. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 17 (6), 669–681. doi:10.1101/gr.6339607
- International Schizophrenia Consortium, Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460 (7256), 748–752. doi:10.1038/nature08185
- Jayasinghe, D., Momin, Md. M., Beckmann, K., Hyppönen, E., Benyamin, B., and Lee, S. H. (2024). Mitigating type 1 error inflation and power loss in GxE PRS: genotype-environment interaction in polygenic risk score models. *Genet. Epidemiol.* 48 (2), 85–100. doi:10.1002/gepi.22546
- Kendler, K. S., and Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* 37 (5), 615–626. doi:10.1017/S0033291706009524
- Lake, R. I. E., Eaves, L. J., Maes, H. H. M., Heath, A. C., and Martin, N. G. (2000). Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45,850 twins and relatives on two continents. *Behav. Genet.* 30 (3), 223–233. doi:10.1023/A:1001918408984
- Lloyd-Jones, L. R., Zeng, J., Sidorenko, J., Yengo, L., Moser, G., Kemper, K. E., et al. (2019). Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* 10 (1), 5086. doi:10.1038/s41467-019-12653-0
- McAdams, T. A., Gregory, A. M., and Eley, T. C. (2013). Genes of experience: explaining the heritability of putative environmental variables through their association with behavioural and emotional traits. *Behav. Genet.* 43 (4), 314–328. doi:10.1007/s10519-013-9591-0
- Morrison, C. L., Rhee, S. H., Smolker, H. R., Corley, R. P., Hewitt, J. K., and Friedman, N. P. (2021). Genetic and environmental influences on stressful life events and their associations with executive functions in young adulthood: a longitudinal twin analysis. *Behav. Genet.* 51 (1), 30–44. doi:10.1007/s10519-020-10017-9
- Neale, M. C., and Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Springer Science and Business Media.
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R., et al. (2016). OpenMx 2.0: extended structural equation and statistical modeling. *Psychometrika* 81 (2), 535–549. doi:10.1007/s11336-014-9435-8
- Ni, G., Zeng, J., Revez, J. A., Wang, Y., Zheng, Z., Ge, T., et al. (2021). A comparison of ten polygenic score methods for psychiatric disorders applied across multiple cohorts. *Biol. Psychiatry* 90 (9), 611–620. doi:10.1016/j.biopsych.2021.04.018
- Palla, L., and Dudbridge, F. (2015). A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am. J. Hum. Genet.* 97 (2), 250–259. doi:10.1016/j.ajhg.2015.06.005
- Privé, F., Arbel, J., and Vilhjálmsdóttir, B. J. (2021). LDpred2: better, faster, stronger. *Bioinformatics* 36 (22–23), 5424–5431. doi:10.1093/bioinformatics/btaa1029
- Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Res. Official J. Int. Soc. Twin Stud.* 5 (6), 554–571. doi:10.1375/136905202762342026
- R Core Team (2025). R: a language and environment for statistical computing. *R Foundation for Statistical Computing*. Available online at: <https://www.r-project.org/>.
- Reed, Z. E., Morris, T. T., Davis, O. S. P., Smith, G. D., Munafò, M. R., and Griffith, G. J. (2022). Examining the association between genetic risk for depression, wellbeing and schizophrenia, and proximity to greenspace. *Epidemiology*. doi:10.1101/2022.04.21.22274122
- Ruan, Y., Lin, Y.-F., Feng, Y.-C. A., Chen, C.-Y., Lam, M., Guo, Z., et al. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* 54 (5), 573–580. doi:10.1038/s41588-022-01054-7
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12 (3), e1001779. doi:10.1371/journal.pmed.1001779
- The All of Us Research Program Investigators, Denny, J. C., Rutter, J. L., Goldstein, D. B., Philippakis, A., Smoller, J. W., et al. (2019). The “All of Us” research program. *N. Engl. J. Med.* 381 (7), 668–676. doi:10.1056/NEJMs1809937
- Truett, K. R., Eaves, L. J., Walters, E. E., Heath, A. C., Hewitt, J. K., Meyer, J. M., et al. (1994). A model system for analysis of family resemblance in extended kinships of twins. *Behav. Genet.* 24 (1), 35–49. doi:10.1007/BF01067927
- Van Der Sluis, S., Posthuma, D., and Dolan, C. V. (2012). A note on false positives and power in G × E modelling of twin data. *Behav. Genet.* 42 (1), 170–186. doi:10.1007/s10519-011-9480-3
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*. 4th Edn. New York, NY: Springer.
- Verhulst, B., and Benstock, S. E. (2023). Toward the next generation of precision medicine. *Policy Insights Behav. Brain Sci.* 10 (2), 256–263. doi:10.1177/23727322231197765

- Verhulst, B., Prom-Wormley, E., Keller, M., Medland, S., and Neale, M. C. (2019). Type I error rates and parameter bias in multivariate behavioral genetic models. *Behav. Genet.* 49 (1), 99–111. doi:10.1007/s10519-018-9942-y
- Verhulst, B., Pritikin, J. N., Clifford, J., and Prom-Wormley, E. (2021). Using genetic marginal effects to study gene-environment interactions with GWAS data. *Behav. Genet.* 51 (3), 358–373. doi:10.1007/s10519-021-10058-8
- Visscher, P. M., Hemani, G., Vinkhuyzen, A. A. E., Chen, G.-B., Lee, S. H., Wray, N. R., et al. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 10 (4), e1004269. doi:10.1371/journal.pgen.1004269
- Werme, J., van der Sluis, S., Posthuma, D., and de Leeuw, C. A. (2021). Genome-wide gene-environment interactions in neuroticism: an exploratory study across 25 environments. *Transl. Psychiatry* 11 (1), 180. doi:10.1038/s41398-021-01288-9
- Xiao, R., and Boehnke, M. (2009). Quantifying and correcting for the winner's curse in genetic association studies. *Genet. Epidemiol.* 33 (5), 453–462. doi:10.1002/gepi.20398
- Zhang, Q., Privé, F., Vilhjálmsson, B., and Speed, D. (2021). Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nat. Commun.* 12 (1), 4192. doi:10.1038/s41467-021-24485-y