

The sample constitutes of 45 3-generations families (unless stated otherwise) with 22 individuals in each family ($n = 990$). Given the covariates and random effects, we generate the binary phenotype according to the following logistic mixed model

$$\begin{aligned} Y_i | \pi_i &\sim \text{Ber}(\pi_i), \\ \text{logit}(\pi_i) &= \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 \mathbf{Z}_i + u_i \\ \mathbf{u} &\sim \text{MVN}(0, \sigma_u^2 \Phi), \end{aligned} \tag{Equation 1}$$

where we include three covariates that affect the trait: age, sex and a $N(0, 1)$ covariate \mathbf{Z} and the intercept term β_0 is chosen to target a specific population prevalence. Random effects \mathbf{u} are included that represent the polygenic additive effects and have covariance matrix Φ that is twice the kinship matrix. We determine our simulation settings based on:

1. $\alpha = \text{Var}(\pi)/\text{Var}(Y)$,
2. $\theta = \sigma_u^2/(\sigma_u^2 + \text{Var}(X_i^T \beta))$,
3. Population prevalence K ,
4. Presence/absence of ascertainment in the sample
5. The number of tests n

We generate sets of rare variants for each phenotype realization that we test for association with the phenotype using Variable Threshold (VT) and Weighted Sum (WS) statistics. In order to assess significance, we generate phenotype replicates based on the estimating equations framework of CARAT as well as through parametric bootstrap.

In the CARAT framework, we convert the quantitative phenotype replicates to binary based on the number of affected in the sample for the WS test (i.e. set the top $n\bar{\mathbf{Y}}$ phenotypic values to 1). (Note: could also use the same procedure but narrow down to each family so that the number of affected within families is the same)

Estimating equations framework of CARAT

No ascertainment

VT	Type 1 error		
Simulation settings	0.05	0.01	0.005
$\alpha = \theta = 0, K = 32\%, n = 22800$	0.0483	0.0087	0.0043
$\alpha = 6\%, \theta = 1, K = 32\%, n = 52500$	0.0528*	0.011	0.0055
$\alpha = 22\%, \theta = 1, K = 35\%, n = 32500$	0.057*	0.011	0.0056

WS	Type 1 error		
Simulation settings	0.05	0.01	0.005
$\alpha = \theta = 0, K = 32\%, n = 22500$	0.0493	0.01	0.0053
$\alpha = 1\%, \theta = 1, K = 32\%, n = 45300$	0.0489	0.0097	0.0057
$\alpha = 3\%, \theta = 1, K = 32\%, n = 45300$	0.0526	0.0104	0.0054
$\alpha = 6\%, \theta = 1, K = 32\%, n = 45300$	0.0553*	0.012*	0.006*

Parametric Bootstrap with MCMCglmm

No ascertainment with unobserved epistasis + prop. of cases is at most 1% different between trait and its replicates for the whole sample. The posterior modes of the fixed effects and VC are used as estimates.

VT	Type 1 error		
Simulation settings	0.05	0.01	0.005
$\alpha = \theta = 0, K = 30\%, n = 29700$	0.047*	0.009	0.0043
$\alpha = 5\%, \theta = .2, K = 30\%, n = 28950$	0.0508	0.011	0.0051
$\alpha = 9\%, \theta = .4, K = 32\%, n = 29850$	0.0528*	0.011*	0.0059*
$\alpha = 13\%, \theta = .6, K = 32\%, n = 30000$	0.0503	0.011	0.0056
$\alpha = 16\%, \theta = .8, K = 32\%, n = 30000$	0.0508	0.0104	0.0053
$\alpha = 20\%, \theta = 1, K = 33\%, n = 28950$	0.0503	0.0095	0.0048
$\alpha = \theta = 0, K = 30\%, n = 29550$	0.0497	0.0107	0.0053
$\alpha = 13\%, \theta = .2, K = 32\%, n = 29850$	0.0484	0.0105	0.0049
$\alpha = 22\%, \theta = .4, K = 34\%, n = 30000$	0.0481	0.0102	0.0049
$\alpha = 30\%, \theta = .6, K = 36\%, n = 30000$	0.047*	0.0085*	0.0041*
$\alpha = 36\%, \theta = .8, K = 37\%, n = 30000$	0.0498	0.0099	0.0049
$\alpha = 41\%, \theta = 1, K = 37\%, n = 29550$	0.0475*	0.0093	0.0041*
$\alpha = \theta = 0, K = 30\%, n = 29550$	0.0482	0.0097	0.0049
$\alpha = 28\%, \theta = .2, K = 35\%, n = 30000$	0.0478	0.009	0.0043
$\alpha = 42\%, \theta = .4, K = 38\%, n = 30000$	0.0487	0.0097	0.0047
$\alpha = 50\%, \theta = .6, K = 40\%, n = 30000$	0.049	0.0092	0.0042
$\alpha = 56\%, \theta = .8, K = 41\%, n = 29700$	0.049	0.0093	0.0049
$\alpha = 60\%, \theta = 1, K = 41\%, n = 29550$	0.0479	0.0088	0.0041

WS	Type 1 error		
Simulation settings	0.05	0.01	0.005
$\alpha = \theta = 0, K = 30\%, n = 29100$	0.0499	0.01	0.0051
$\alpha = 5\%, \theta = .2, K = 30\%, n = 28050$	0.0517	0.0099	0.0052
$\alpha = 9\%, \theta = .4, K = 32\%, n = 29100$	0.0512	0.0111	0.0054
$\alpha = 13\%, \theta = .6, K = 32\%, n = 29250$	0.0521	0.0099	0.0045
$\alpha = 16\%, \theta = .8, K = 32\%, n = 29700$	0.0485	0.0102	0.0051
$\alpha = 20\%, \theta = 1, K = 33\%, n = 29850$	0.0499	0.0093	0.0048
$\alpha = \theta = 0, K = 30\%, n = 29100$	0.0479	0.0086*	0.0048
$\alpha = 13\%, \theta = .2, K = 32\%, n = 29850$	0.0497	0.0096	0.0045
$\alpha = 22\%, \theta = .4, K = 34\%, n = 30000$	0.0452*	0.0082*	0.0038*
$\alpha = 30\%, \theta = .6, K = 36\%, n = 29850$	0.0485	0.0095	0.0047
$\alpha = 36\%, \theta = .8, K = 37\%, n = 29700$	0.0487	0.0091	0.0044
$\alpha = 41\%, \theta = 1, K = 37\%, n = 29250$	0.0486	0.0092	0.0044
$\alpha = \theta = 0, K = 30\%, n = 29250$	0.048	0.009	0.0048
$\alpha = 28\%, \theta = .2, K = 35\%, n = 30000$	0.0494	0.0098	0.0045
$\alpha = 42\%, \theta = .4, K = 38\%, n = 30000$	0.0489	0.0101	0.0047
$\alpha = 50\%, \theta = .6, K = 40\%, n = 29700$	0.0506	0.0087*	0.0043
$\alpha = 56\%, \theta = .8, K = 41\%, n = 29250$	0.0488	0.0095	0.0046
$\alpha = 60\%, \theta = 1, K = 41\%, n = 29700$	0.049	0.0092	0.0045

With ascertainment (6-8 affected) and unobserved epistasis + prop. of cases is at most 1% different between trait and its replicates for the whole sample. The posterior modes of the fixed effects and VC are used as estimates.

VT	Type 1 error		
Simulation settings	0.05	0.01	0.005
$\alpha = 9\%, \theta = .4, K = 32\%, n = 29850$	0.0398	0.007	0.0038*
$\alpha = 16\%, \theta = .8, K = 32\%, n = 30000$	0.0463	0.0078	0.0034

WS	Type 1 error		
$\alpha = 5\%, \theta = .4, K = 12\%, n = 29100$	0.0373	0.0064	0.0031
$\alpha = 12\%, \theta = .8, K = 13\%, n = 29700$	0.0383	0.0070	0.0034