



VCU

Virginia Commonwealth University
VCU Scholars Compass

Graduate Research Posters

Graduate School

2021

PRINCIPAL COMPONENTS ANALYSIS CORRECTS COLLIDER BIAS IN POLYGENIC RISK SCORE EFFECT SIZE ESTIMATION

Nathaniel S. Thomas
Virginia Commonwealth University

Peter B. Barr
Virginia Commonwealth University

Fazil Aliev
Virginia Commonwealth University

See next page for additional authors

Follow this and additional works at: <https://scholarscompass.vcu.edu/gradposters>



Part of the [Biostatistics Commons](#), [Molecular Genetics Commons](#), and the [Psychology Commons](#)

Downloaded from

Thomas, Nathaniel S.; Barr, Peter B.; Aliev, Fazil; Kuo, Sally I.; Dick, Danielle M.; and Salvatore, Jessica E., "PRINCIPAL COMPONENTS ANALYSIS CORRECTS COLLIDER BIAS IN POLYGENIC RISK SCORE EFFECT SIZE ESTIMATION" (2021). *Graduate Research Posters*. Poster 115.
<https://scholarscompass.vcu.edu/gradposters/115>

This Poster is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Graduate Research Posters by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

Authors

Nathaniel S. Thomas, Peter B. Barr, Fazil Aliev, Sally I. Kuo, Danielle M. Dick, and Jessica E. Salvatore



VCU



Principal Components Analysis Corrects Collider Bias in Polygenic Risk Score Effect Size Estimation

Nathaniel S. Thomas, Peter B. Barr, Fazil Aliev, Sally I. Kuo., Danielle M. Dick, Jessica E. Salvatore

Background

- Genome-wide polygenic scoring has emerged as a way to predict psychiatric and behavioral outcomes and identify environments that promote the expression of genetic risks.
- The effects of polygenic risk scores (PRS) may be biased by the inclusion of heritable environments as covariates when the environment is influenced by unmeasured confounding variables, an example of collider bias.
- For example, the estimated effect of a polygenic score for alcohol consumption may be biased in a model that also includes educational attainment as a covariate. Educational attainment is heritable and likely to be influenced by a wide variety of other unmeasured variables
- Inclusion of the principal components of observed confounders as covariates may correct for the effect of unmeasured confounders.

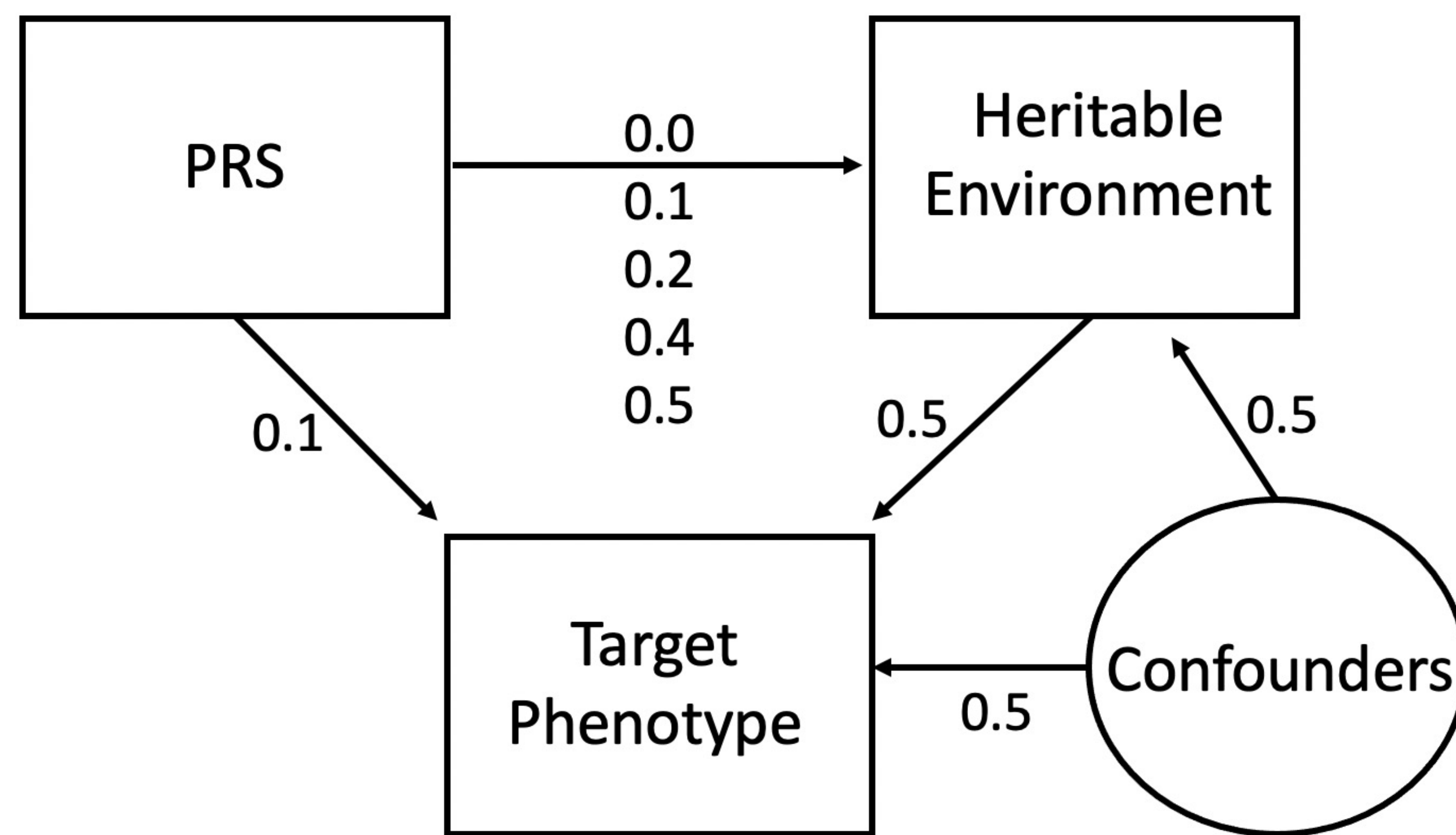
Method

- A simulation study was conducted to test principal components analysis (PCA) as a correction for collider bias.
- Data were sampled from a model which tested different values for...
 - The effect of the polygenic risk score on the heritable environment (rGE)
 - Path values fixed to: 0, 0.1, 0.2, 0.4, 0.5
 - The correlation structure of the unmeasured confounding data
 - Correlations drawn from a uniform distribution with ranges: 0.05 – 0.1, 0.2 – 0.3, 0.5 – 0.6, 0.8 – 0.9
 - The proportion of the confounding data that is used to construct the principal component
 - Varying from 3% (3 variables) to 100% (100 variables)
- Other model parameters were fixed across all simulation iterations, as shown in Figure 1

Results

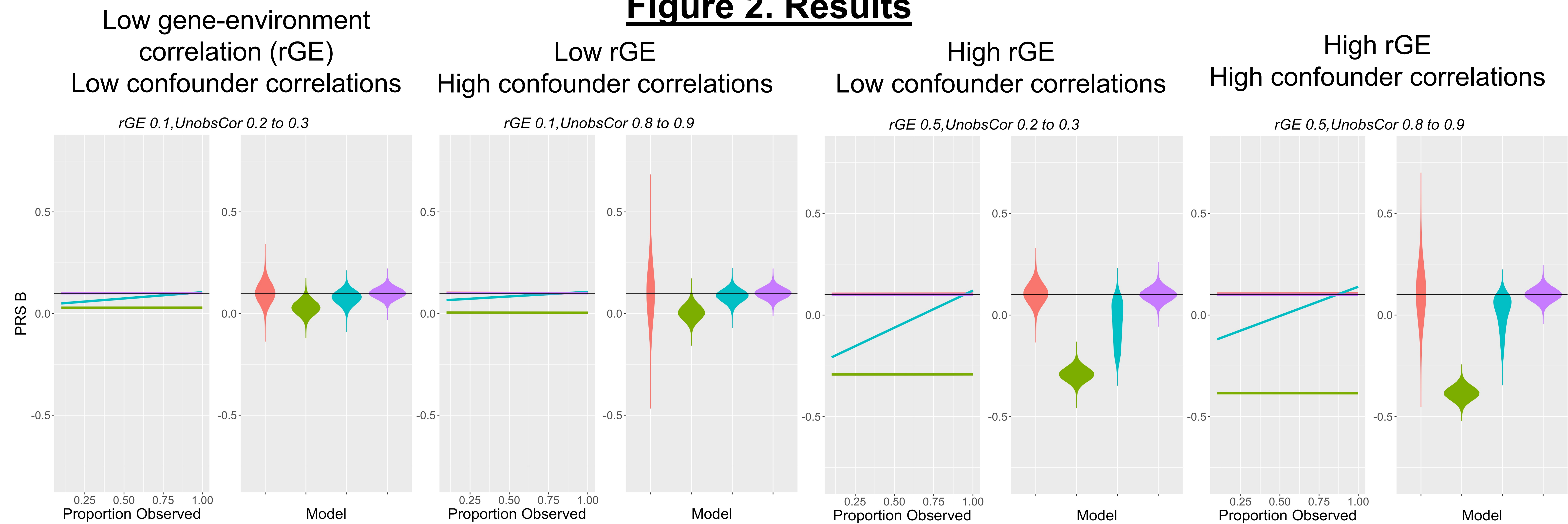
- Modeling the first PC of observed confounders as a covariate recovers the PRS effect size estimate under reasonable assumptions about...
 - The proportion of the confounding data that is measured
 - The correlation structure of the confounding data
- Required assumptions become stricter as the effect of PRS on environment (and the magnitude of bias) increases.
- Figure 2 depicts results for a sample of the tested parameters that are representative of the general trends across all versions of the simulation

Figure 1. Simulation Model



Controlling for the first PC of confounding data corrects collider bias, improving the accuracy of PRS effect size estimation in the presence of heritable environmental covariates.

Figure 2. Results



Red: Univariate PRS B estimate (Phenotype ~ PRS)

Green: Collider Bias PRS Estimate (Phenotype ~ PRS + Environment)

Blue: Corrected PRS Estimate, at varying levels of % confounders observed (Y ~ PRS+ Environment + subset data PC)

Purple: Fully Corrected PRS Estimate, using all confounders in PCA (Y ~ PRS+ Environment + full data PC)

