

NAME OF PRESENTING AUTHOR: Nathaniel S. Thomas

EMAIL ADDRESS OF PRESENTING AUTHOR: thomasns@vcu.edu

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

Nathaniel S. Thomas, M.S.¹, Peter Barr, Ph.D.¹, Fazil Aliev, Ph.D.¹, Mallory Stephenson, M.S.³, Sally I. Kuo, Ph.D.¹, Danielle M. Dick, Ph.D.^{1,2}, Jessica E. Salvatore, Ph.D.^{1,3}

¹ Department of Psychology, Virginia Commonwealth University, Box 842018, Richmond, VA 23284-2018.

² Department of Human & Molecular Genetics, Virginia Commonwealth University, Box 980033, Richmond, VA 23298-0033.

³ Virginia Institute for Psychiatric and Behavioral Genetics, Box 980126, Richmond, VA 23298-0126.

KEYWORDS: Polygenic Risk Scores; Collider Bias; Principal Components Analysis

ABSTRACT:

Genome-wide polygenic scoring has emerged as a way to predict psychiatric and behavioral outcomes and identify environments that moderate the expression of genetic predispositions. An increasing number of studies demonstrate that the inclusion of heritable environments as covariates may bias the effects of polygenic risk scores (PRS) when the environmental covariates are influenced by unmeasured confounding variables, an example of collider bias. We tested the hypothesis that the inclusion of the principal components of observed confounders as covariates may correct for the effect of unmeasured confounders. We conducted a simulation study to test principal components analysis (PCA) as a correction for collider bias. Data were sampled from a model that tested different values for the effect of the polygenic risk score on the heritable environment, the correlation structure of the unmeasured confounding data, and the proportion of the confounding data that is used to construct the principal components. Other model parameters were fixed across all simulation iterations. 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 or the correlation structure of the confounding data. Required assumptions become stricter as the association between PRS and the heritable environment (and the magnitude of bias) increases. Inclusion of the first PC of observed confounders may improve the accuracy of PRS effect size estimation when heritable environments are included in the model as covariates. Future directions include application of this method in observed data.

GRANT SUPPORT: R01AA028064