

Journal club:

Adjusting for heritable covariates can bias effect in genome-wide association studies

Hugues Aschard et al

The American Journal of Human Genetics

February 5, 2015

Overview

1 Introduction

- Context
- Purpose
- Underlying Causal Diagrams
- Model
- Effect Estimates and False Discovery Rate
- Potential presence of bias due to adjustment

2 Illustration

- Theory
- Simulation
- Heritability of Adjusted Phenotypes

3 Conclusion

Context

Heritable human traits have genetic associations.

Lots of large-scale GWASs published for human traits are adjusted for other correlated traits with a genetic basis.

Theoretically:

Genetic variant can be associated with both the primary outcome and the covariate used for adjustment.

So:

The adjusted and unadjusted estimated effects of the genetic variant on the outcome will differ.

Motivation: discover genetic variants associated with the primary outcome independently of the correlated trait.

Purpose

Purposes adjustment of covariates in GWASs:

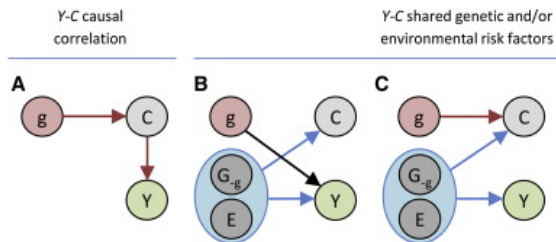
- account for potential confounding factors that can bias SNP effect estimates
- improve statistical power by reducing residual variance

Examples: to account for population structure, correlated environmental or demographic factors, to capture batch effects in gene-expression analysis.

⇒ To increase statistical power.

Underlying Causal Diagrams

- 1 genotypes G ,
- 2 environment E ,
- 3 covariate C ,
- 4 outcome of interest Y .

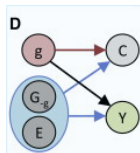


The strength of this association (Fig1C) depends on:

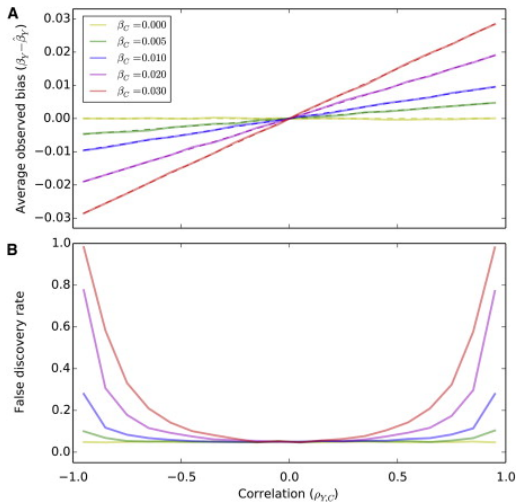
- ρ_{CY} = Correlation between covariate and outcome due to shared risk factors,
- β_C = Effect of the genetic variant on the covariate,
- $\widehat{\beta}_Y$ = Bias of the genetic effect estimate (for normalized g, C, and Y with mean 0 and variance 1)
 $\simeq -\beta_C \rho_{CY}$ (if β_C small and sample size sufficiently large).

Effect Estimates and False Discovery Rate

Product between the direct genetic effect estimate on the covariate and the correlation between the outcome and the covariate.



Increased false
discovery rates under
the null.



Potential presence of bias due to adjustment

Adjusting for covariate with no genetic component

→ \emptyset bias for genotype effect ($\beta_C = 0$).

Adjusting for covariate with genetic component ($\beta_C \neq 0$)

→ Adjusted association signals difficult to interpret:

- can correspond also to a bivariate signal on Y and C,
- or to an association with the covariate only.

Illustration

Examples from published GWAS, 23 SNPs in gender specific samples associated with:

waist-to-hip ratio(WHR) and waist circumference(WC) adjusted for body mass index (BMI) by Heid et al. and Randal et al.

Effects on WHR or WC (before/after adjustment for BMI) + Effect BMI
→ extracted from the GIANT (statistics database)

Hypothesis: Genetic effects might be biased as a result of adjustment for body mass index.

Theory

Theory 1

The adjusted statistical test can have **increased power** to detect the genetic variant, as compared to the unadjusted test, if the genetic effect and the phenotypic correlation are in **opposite directions**.

Theory 2

Conversely, if the genetic effect and the correlation are in the **same direction**, the adjusted statistical test has, in many cases, a **decreased power** to detect the genetic variant

SNPs harboring opposite marginal effects on the two traits are significantly enriched ($p = 0.005$)

Hypothesis: Absence of genetic effect on BMI \rightarrow # SNP with opposite directions of effect follows a binomial distribution ($p = 0,5$).

Enrichment observed: A substantial fraction of these SNPs is associated with BMI in the opposite direction.

\Rightarrow Even non-significant genetic effects on the covariate can influence power when correlation between the outcome and the covariate is large (e.g., 0.5).

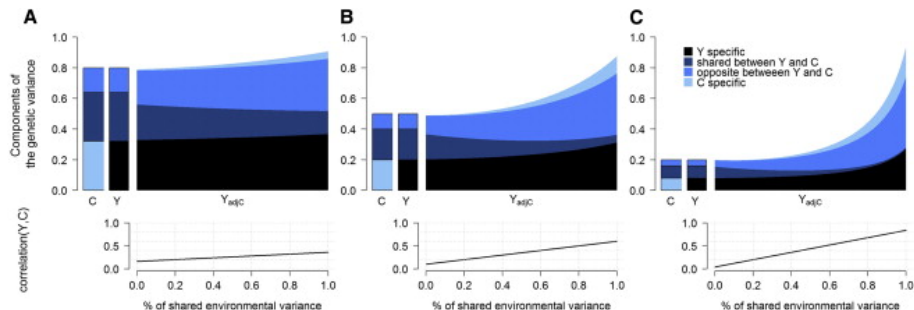
Heritability of Adjusted Phenotypes

Heritability of a given phenotype VS. Heritability adjusted correlated variable.

Upper panel: Heritability variation of Y and C :

- 0.8 (A), 0.5 (B), 0.2 (C)

Bottom panel: Proportion of shared environmental variance from 0 to 1.



Conclusion

If purpose of an adjusted analysis is:

- to increase statistical power rather than detect unbiased direct effects,
→use multivariate approaches (powerful to detect pleiotropic loci affecting multiple traits)
- if adjusted analyzes are performed,
→report estimates of genetic effects on the covariate and pre- and post-adjustment outcomes, their standard deviation and significance, and the correlation between outcome and covariate.

⇒ The magnitude of a potential bias can be estimated and taken into account when interpreting the results.

Thank you for your attention