

## Response to Day et al.

*To the Editor:* Day et al. present a striking example of how adjusting for heritable covariates correlated with the outcome in a genetic association study can bias genetic effect estimates—possibly creating a strong association where genotype has no causal effect on outcome. This bias was the focus of our recent report.<sup>1</sup> As Day et al. point out, this phenomenon is a special case of the broader concept of collider-stratification bias<sup>2,3</sup>. By choosing an outcome that is not associated with autosomal variation (gender), but strongly associated with the covariate (height), Day et al. demonstrate that collider-stratification bias can lead to many very strong false-positive signals. Their results also nicely confirm the direct relationship between gene-covariate effect estimates and gene-outcome effect estimates after adjustment for the covariate. The most significant p value for rs724016 after adjustment ( $P_{\text{adj}} = 7 \times 10^{-90}$ ) is impressive but actually expected—even without running the analysis. The most recent meta-analysis genome-wide association study (GWAS) of height reported a Z score of  $-26.9$  ( $p = 3.2 \times 10^{-158}$ ) for association between rs724016 and height for a sample size of 252,972.<sup>4</sup> Assuming that the effect of the SNP on height is similar across the UK Biobank and GIANT (genetic investigation of anthropometric traits) consortium participants, and assuming a correlation between height and gender of 0.71, as derived from the 2013–2014 NHANES (National Health and Nutrition Examination Survey) survey for non-Hispanic white individuals over 20 years old,<sup>5</sup> the expected chi-square in a height-adjusted analysis under the null hypothesis is  $\chi^2 \sim (-26.9 \times 0.71)^2 \times (142,630/252,972) / (1 - 0.71^2) = 414.6$ . Similarly to the Day et al. experiment, this corresponds to a highly significant p value of  $3 \times 10^{-92}$ . Interestingly, the height-sex correlation is almost equal to the value where the significance of the observed signal in the adjusted analysis matches the significance of the SNP-covariate association test, assuming the same sample size in both analyses. For a stronger correlation, the adjusted analysis would result in a more significant association on average. Then, looking at the SNP-covariate association p value would no longer be a good indicator of potential bias, presenting a worrisome situation.

Although the example of Day et al. confirms the potential for collider bias, their example assumes that the underlying causal model is understood so that the genetic effect of the variants can be safely interpreted. However, in practice, the underlying mechanism is generally unknown, or at least subject to debate, and for some investigators the question of whether to adjust for a

covariate remains. Taking height as an example, it is intuitive to adjust for height when considering phenotypes such as pulmonary function, bone mineral density, intracranial volume, and body measurements in general, given that they are proportional to height. The common argument for the adjustment has been about focusing on factors associated with the outcome independently of height—in other words, factors altering this proportionality. The adjustment will partly fulfill this goal by (1) reducing the signal at SNPs harboring positive pleiotropy (i.e., SNPs positively associated with both height and the outcome) and (2) enhancing the signal for variants associated with the primary outcome only and those with negative pleiotropic effects. However, because of height polygenicity, it will also (3) induce a false signal at all genetic variants associated with height only (as observed in the experiment by Day et al.). Estimation of the shared genetic correlation between the outcome and the covariate, e.g., as done by Bulik-Sullivan et al.,<sup>6</sup> can provide a first approximation of the proportion of SNPs falling into scenarios one, two, and three described above, and thus can help with the decision of whether or not to adjust.

Hugues Aschard,<sup>1,2,\*</sup> Bjarni J. Vilhjálmsson,<sup>3</sup>  
Amit D. Joshi,<sup>1,2,4</sup> Alkes L. Price,<sup>1,2</sup> and Peter Kraft<sup>1,2</sup>

<sup>1</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA; <sup>2</sup>Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA; <sup>3</sup>Bioinformatics Research Centre, Aarhus University, Aarhus 8000, Denmark; <sup>4</sup>Division of Gastroenterology, Massachusetts General Hospital, Boston, MA 02114, USA

\*Correspondence: [haschard@hsph.harvard.edu](mailto:haschard@hsph.harvard.edu)

## References

- Aschard, H., Vilhjálmsson, B.J., Joshi, A.D., Price, A.L., and Kraft, P. (2015). Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* 96, 329–339.
- Greenland, S. (2003). Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 14, 300–306.
- Greenland, S., Pearl, J., and Robins, J.M. (1999). Causal diagrams for epidemiologic research. *Epidemiology* 10, 37–48.
- Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46, 1173–1186.

5. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data, 2013-2014. [http://www.cdc.gov/nchs/nhanes/nhanes\\_questionnaires.htm](http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm).
6. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al.; ReproGen Consortium; Psychiatric Genomics Consortium; Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3 (2015). An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* *47*, 1236–1241.

<http://dx.doi.org/10.1016/j.ajhg.2015.12.020>. ©2016 by The American Society of Human Genetics. All rights reserved.