

Genetic instruments with too many strings: acknowledging pleiotropy and population structure in Mendelian randomization studies

Jeremy A. Labrecque ^{1*}, Jay S. Kaufman ², and Jonathan Y. Huang ³

¹Department of Epidemiology, Erasmus MC, Dr. Molewaterplein 50, 3015 GE Rotterdam, the Netherlands; ²Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Avenue des Pins Ouest, H3A 1A2 Montreal, Quebec, Canada; and ³Department of Biostatistics, Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research (A*STAR), 117609, Singapore

Online publish-ahead-of-print 20 January 2020

This commentary refers to ‘Genetically modulated educational attainment and coronary disease risk’, by L. Zeng et al., pages 2413–2420.

The availability of large genetic databases has prompted renewed interest in genetic contributions to educational attainment and the effects of education on health. In a recent issue of the *European Heart Journal*, Zeng et al.¹ use cohorts from across Europe and the USA to investigate education-associated genotypes and their role in coronary artery disease (CAD). We would like to commend the authors for raising important questions about the relationships between genetics, education, and CAD. Nonetheless, it is our belief that the methods employed have important limitations not discussed by the authors. Here, we note three of these limitations.

First, the authors investigate whether the genetic risk score (GRS) for education was associated with CAD independently of its effect on education by adjusting for number of years of attained education in the regression of CAD on the GRS. The logic is to ‘block’ the effect of the GRS through attained education (Figure 1). However, beyond ‘blocking’ the effect, this method will also create a new correlation between GRS and all other causes of education because it becomes a ‘collider’² (Figure 1, dashed arrow). Therefore, even if the effect of the education GRS on CAD is only mediated by years of education, the GRS will continue to be correlated with CAD after adjustment for years of education. Any residual association after adjusting for years of education cannot then be interpreted as an independent or pleiotropic effect of the GRS. To answer this question by adjustment, appropriate mediation methods would be required.³

Second, one of the three major assumptions of Mendelian randomization (MR) analyses is that the effect of the single-nucleotide polymorphisms on the outcome (CAD) only passes through the exposure (education).⁴ Therefore, if the authors’ claim to have found pleiotropic effects of the education GRS on CAD were, in fact, true,

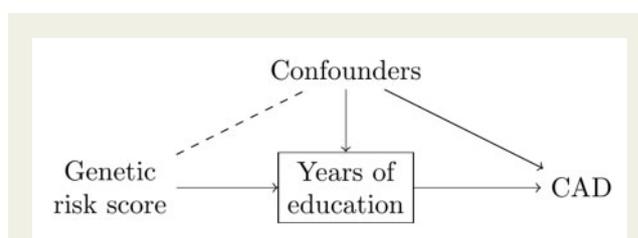


Figure 1 Adjusting for years of education blocks the effect of the genetic risk score on coronary artery disease but creates a correlation between the genetic risk score and the confounders of education and coronary artery disease (dashed line).

this would also imply that the results of the MR analysis are biased. Therefore, it is difficult to know what to conclude. If the education GRS is pleiotropic, as the authors claim, then the main MR analysis must be biased. Alternatively, if the authors claim that the MR analysis is valid, this implies that there is no pleiotropic effect of the education GRS on CAD.

Lastly, confounding by population stratification is difficult to control when using a GRS, particularly, with geographically patterned phenotypes, as in the case of education and heart disease.⁵ The authors use five principal components to adjust for this but it has been shown that geographic confounding may not be successfully controlled even with 40 principal components. This confounding would invalidate both the primary and MR analyses. Furthermore, such residual confounding is consistent with the finding that GRS-CAD associations become null when adjusting for body mass index and smoking, i.e. adjusting for these factors help account for population stratification thus showing the independent association with GRS to be null.

* Corresponding author. Tel: +31 10 70 33638, Email: j.labrecque@erasmusmc.nl

Published on behalf of the European Society of Cardiology. All rights reserved. © The Author(s) 2020. For permissions, please email: journals.permissions@oup.com.

Conflict of interest: none declared.

References

1. Zeng L, Ntalla I, Kessler T, Kastrati A, Erdmann J; UK Biobank CardioMetabolic Consortium CHD Working Group, Danesh J, Watkins H, Samani NJ, Deloukas P, Schunkert H. Genetically modulated educational attainment and coronary disease risk. *Eur Heart J* 2019;**40**:2413–2420.
2. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C. Illustrating bias due to conditioning on a collider. *Int J Epidemiol* 2010;**39**:417–420.
3. Vanderweele TJ. *Explanation in Causal Inference*. New York, NY: Oxford University Press; 2015.
4. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
5. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, Carslake D, Hemani G, Paternoster L, Smith GD, Davies N, Lawson DJ, J Timpson N. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun* 2019;**10**: 333.