Problems in interpreting and using GWAS of conditional phenotypes illustrated by "alcohol GWAS"

Michael V Holmes[1-4] MD PhD and George Davey Smith[4-6] MD, DSc

[1] Medical Research Council Population Health Research Unit at the University of Oxford, UK.
[2] Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Big Data Institute Building, Old Road Campus, Roosevelt Drive, Oxford OX3 7LF, UK.
[3] National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospital, Oxford, UK
[4] Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom.
[5] School of Social and Community Medicine, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom.
[6] National Institute for Health Research Bristol Biomedical Research Centre, Oakfield House, Oakfield Grove, Bristol BS8 2BN

Correspondence to

Michael Holmes, Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, UK michael.holmes@ndph.ox.ac.uk

and

George Davey Smith, Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK KZ.Davey-Smith@bristol.ac.uk

Conflicts of Interest: The authors declare no competing interests in relation to the work described.

Dear Editor,

We read with interest findings of a genome-wide association study (GWAS) in the UK Biobank reported by Clarke et al[1] in *Molecular Psychiatry*. The authors aimed to conduct a GWAS of alcohol consumption. Prior to conducting the GWAS, Clarke and colleagues regressed alcohol intake on body weight and age, and used the residuals from this regression analysis for the GWAS analysis. The authors identified 14 loci harbouring variants associated with the residuals at levels indicative of GWAS significance and took these forward to characterise the relationship of these SNPs with various phenotypes, such as adiposity.

By adjusting alcohol for body weight prior to the GWAS, Clarke and colleagues in effect generate a phenotype which is automatically associated with body weight and its correlates.[2] When the SNPs from the GWAS are combined (as the authors do), this would have the effect of generating a genetic instrument that has a positive association with alcohol and an inverse association with body weight. This can lead to spurious findings when the SNPs are then related to other traits. For example, in figure 3, the authors note a negative genetic correlation of alcohol consumption (indexed by SNPs associating with alcohol adjusted for body weight) with various markers of adiposity including waist circumference, body mass index, overweight and obesity. In Table 2, these relationships are more formally characterised: a polygenic risk score associating with 0.08 SD higher alcohol per week was associated with a 0.024 SD lower body mass index (BMI). A naïve interpretation would be that more alcohol leads to less adiposity, however this would be incorrect. Large-scale Mendelian randomization meta-analysis using rs1229984 in *ADH1B* has identified the genetic variant associated with more alcohol intake to be strongly associated with a higher (not lower) body mass index.[3] A recent instrumental variable study that used flushing as a proxy for *ALDH2* genotype also suggested strong associations of more alcohol with higher BMI [4]. Thus, a large body of evidence using reliable Mendelian randomization approaches indicates that consuming more alcohol, all things being equal, leads to higher BMI.

So why the discordance between these Mendelian randomization studies and the current report by Clarke and colleagues? As described above, by regressing alcohol for body weight prior to the GWAS, this means that SNPs that were identified by the GWAS had a pattern of association indicative of higher alcohol but lower body weight. In effect, the associations that Clarke and colleagues report with adiposity have been induced by the very nature of their statistical model.

An analogy is conducting a GWAS for waist hip ratio (WHR) adjusted for BMI (WHRadjBMI) – on combining the 49 WHRadjBMI SNPs together, the SNPs collectively associate with higher WHR but also associate with lower BMI[5, 6]. This means that while the aim of deriving the

WHRadjBMI phenotype was to proxy central adiposity independently of general adiposity, the net consequence is to derive a genetic instrument that has opposing effects of contrasting adiposity elements, which can introduce confounding[7] and make challenging the interpretation of findings when taking these variants forward into the Mendelian randomization setting.

The findings by Clarke et al[1] highlight the complexities that can arise in Mendelian randomization analysis[8]. We caution against interpreting these data as suggesting that alcohol consumption is correlated with "many positive health and behavioural traits". The burden of evidence indicates that higher alcohol consumption leads to higher levels of many deleterious cardiovascular risk factors including adiposity, blood pressure, inflammation markers, coronary calcification and vascular disease.

References

1. Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112,117). *Mol Psychiatry* 2017.

2. Aschard H, Vilhjalmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *American journal of human genetics* 2015; **96**(2)**:** 329-339.

3. Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z *et al.* Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *Bmj* 2014; **349:** g4164.

4. Yun KE, Chang Y, Yun SC, Davey Smith G, Ryu S, Cho SI *et al.* Alcohol and coronary artery calcification: an investigation using alcohol flushing as an instrumental variable. *Int J Epidemiol* 2017.

5. Dale C, Fatemifar G, Palmer T, White J, Prieto-Merino D, Zabaneh D *et al.* Causal associations of adiposity and body fat distribution with coronary heart disease, stroke subtypes and type 2 diabetes: a Mendelian randomization analysis. *Circulation* 2017; **In Press**.

6. Emdin CA, Khera AV, Natarajan P, Klarin D, Zekavat SM, Hsiao AJ *et al.* Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *Jama* 2017; **317**(6)**:** 626-634.

7. Davey Smith G, Paternoster L, Relton C. When Will Mendelian Randomization Become Relevant for Clinical Practice and Public Health? *Jama* 2017; **317**(6)**:** 589-591.

8. Holmes MV, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017.