



# Commentary: Tobacco consumption and body weight: Mendelian randomization across a range of exposure

Frank Dudbridge

Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: frank.dudbridge@lshtm.ac.uk

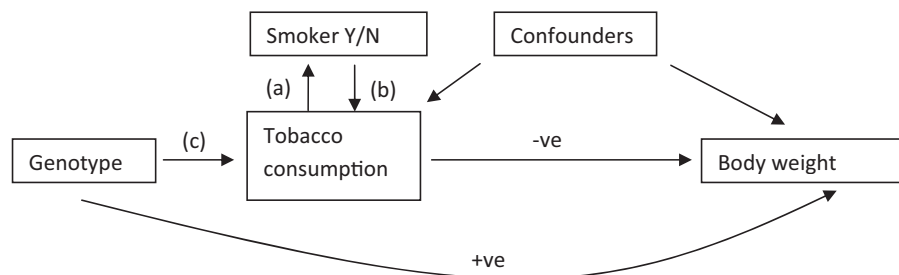
Tobacco consumption is consistently associated with reduced body weight, creating an incentive to initiate smoking and a disincentive to cease, although the health risks associated with the habit outweigh the benefits of reduced weight. Among smokers however, increasing consumption has been associated with increased body weight. To determine whether this contradiction reflects causal processes, Winsløw *et al.*<sup>1</sup> have applied Mendelian randomization (MR) in testing the association of a genetic variant, rs1051730 in *CHRNA3*, with measures of body weight among 80 342 members of the Copenhagen General Population Study. Among smokers, each minor (T) allele carried was associated with an increase of about one cigarette per day, but with a decrease in several measures of body weight, in contrast to the observational results. These results, in line with other recent studies,<sup>2–6</sup> suggest that increased tobacco consumption causes reduced body weight, as does smoking itself.

Here I remark on two aspects of this study that may recur in other MR studies of this type: the restriction of genetic effects to current smokers, and the change, in the observational data, from decreasing to increasing body weight as cigarette consumption increases.

The associations of rs1051730 with decreased body weight are present in current smokers, but not in former or never smokers, suggesting that the gene acts on body weight only through its effect on smoking. However, by stratifying on smoking status, the results are potentially prone to collider bias, whereby an association is induced between the gene and confounders of the exposure–outcome association, creating a non-causal association between the gene and the outcome. The question is whether

smoking status should be considered as derived from the quantity consumed (one who smokes zero cigarettes per day is a non-smoker, others are smokers, **Figure 1** arrow a), or as an exogenous variable whose value constrains the possible consumption (non-smokers must smoke no cigarettes per day, smokers must smoke a positive number, **Figure 1** arrow b). Under the former, stratifying on smoking status would entail a collider bias, and an association would be seen between genotype and smoking status; but such an association was not observed by Winsløw *et al.*, nor was any association between rs1051730 and known confounders when stratifying on smoking status. Thus the data support the view that smoking status is distinct (i.e. has distinct determinants) from tobacco consumption (**Figure 1** arrow b). Indeed it seems reasonable that individuals would generally decide whether or not to smoke as a precursor to developing their usual consumption. Nevertheless, collider bias can only be completely ruled out if the whole sample is analysed, and here the genetic associations remained significant owing to the sufficiently high proportion of smokers among them.

The lack of genetic association among never smokers could be interpreted as evidence against alternative pathways to that through smoking, in line with MR assumptions. However, Taylor *et al.*<sup>6</sup> reported positive association between rs1051730 and body mass index (BMI) ( $P = 6.4 \times 10^{-5}$ ) in a sample of 66 809 never smokers. This trend is apparent in the 32 937 never smokers in Winsløw *et al.* ( $P = 0.07$ , their Figure 4) and would have reached significance ( $P = 0.01$ ) if the same trend were observed in 66 809 subjects. Thus the exclusion restriction seems to be violated, but as the effect is in the opposite direction to



**Figure 1.** Directed acyclic graph showing causal relationships between genotype, smoking status and consumption, and body weight outcomes. Arrow (a) applies if smoking status is defined by the quantity consumed. Arrow (b) applies if smoking status has distinct determinants and defines the possible range of consumption. Arrow (c) is absent for non-smokers under (b). Y/N, yes or no; +ve, positive; -ve, negative.

that observed in smokers, it need not conflict with the qualitative conclusion that increased tobacco consumption reduces body weight.

Observational data show a decrease in body weight at low tobacco consumption, with increases in body weight at higher levels of consumption (Figure 2 of Winslow *et al.*). This raises the question of whether the J-shaped pattern reflects causal effects: the marginal inverse association of rs1051730 may be driven by effects at certain consumption levels. Perhaps there is a particularly strong inverse genetic association at low consumption levels, which outweighs positive genetic associations at higher levels. The conclusion that increased smoking reduces weight might only be applicable at certain low levels of smoking.

Similar questions have recently been addressed in relation to alcohol consumption. rs1229984 in *ADH1B* is associated with reduced alcohol intake and also with reduced incidence of coronary heart disease (CHD) among drinkers, but is not associated with CHD among non-drinkers.<sup>7</sup> This suggests beneficial effects of reduced intake, but observational data show a J-shape with increased risk in abstainers compared with moderate drinkers. Furthermore, for alcohol it is more plausible that abstinence is a special case of low intake, perhaps due to unpleasant side effects associated with *ADH1B* variation. rs1229984 is indeed associated with drinking status,<sup>7</sup> so stratification on drinking status is less defensible than it is for smoking status.

To address these questions, methods have recently been proposed to estimate localised causal effects that can indicate whether the direction of causal effect is constant over the range of exposure. A foundational result is that the marginal treatment effect is a weighted sum of local average treatment effects (LATEs), defined for subgroups of subjects with the same potential outcomes.<sup>8</sup> Silverwood *et al.*<sup>9</sup> considered the effect of alcohol intake on markers of cardiovascular disease. Using rs1229984 in *ADH1B*, they stratified subjects according to their potential alcohol intake if homozygous for the common allele. With some

assumptions their approach explicitly calculates LATEs, and was used to infer J-shaped causal associations of alcohol intake on several cardiovascular markers including systolic blood pressure and BMI. Burgess *et al.*<sup>10</sup> proposed a more general approach, also based on considering potential levels of exposure for a set genetic value. Because the ‘instrument-free exposure’ is by construction not associated with the genotype, stratification on it does not induce a collider bias. With arbitrary stratification, their approach estimates localized causal effects, distinct from formal LATEs but also able to identify non-linear patterns in the causal dose-response curve.

These approaches show promise for developing insight into the pattern of causal effects when observational data show variation in effect size or direction across the range of exposure. One limitation is that the localized effects apply to levels of the instrument-free exposure rather than the exposure itself, and this distorts their interpretation.<sup>11</sup> Thus we can only make statements about subjects with given potential levels of exposure, rather than their actual exposures which are arguably more relevant. Another limitation is an assumption that the effect of genotype on exposure is constant for all subjects. This will not hold when there is a gene-environment interaction, as in the present case in which the gene only affects tobacco consumption in those who are smokers (Figure 1 arrow c); stratification on exposure levels may be required, with attendant possibilities of collider bias.

An increasing number of MR studies will consider dosage of exposures that are present among some subjects only. Such studies must carefully consider the relationship between presence and dosage of exposure, together with their genetic associations. When observational effects vary over the range of exposure, it will be informative to apply emerging methods for localized effects to infer non-linear causal effects. Such approaches will be particularly valuable when there are conflicts between causal and observational results, which merit exploring in greater detail.

## Funding

This work was funded by the MRC (K006215).

**Conflict of interest:** None declared.

## References

1. Winsløw UC, Rode L, Nordestgaard BG. High tobacco consumption lowers body weight: a Mendelian randomization study of the Copenhagen General Population Study. *Int J Epidemiol* 2015; doi:10.1093/ije/dyu276.
2. Freathy RM, Kazeem GR, Morris RW *et al*. Genetic variation at CHRNA5-CHRNA3-CHRNA4 interacts with smoking status to influence body mass index. *Int J Epidemiol* 2011;40:1617–28.
3. Asvold BO, Bjorngaard JH, Carslake D *et al*. Causal associations of tobacco smoking with cardiovascular risk factors: a Mendelian randomization analysis of the HUNT Study in Norway. *Int J Epidemiol* 2014;43:1458–70.
4. Kaakinen M, Ducci F, Sillanpaa MJ, Laara E, Jarvelin MR. Associations between variation in CHRNA5-CHRNA3-CHRNA4, body mass index and blood pressure in the Northern Finland Birth Cohort 1966. *PLoS One* 2012;7:e46557.
5. Varga TV, Hallmans G, Hu FB, Renstrom F, Franks PW. Smoking status, snus use, and variation at the CHRNA5-CHRNA3-CHRNA4 locus in relation to obesity: the GLACIER study. *Am J Epidemiol* 2013;178:31–37.
6. Taylor AE, Morris RW, Fluharty ME *et al*. Stratification by smoking status reveals an association of CHRNA5-A3-B4 genotype with body mass index in never smokers. *PLoS Genet* 2014; 10:e1004799.
7. Holmes MV, Dale CE, Zuccolo L *et al*. Association between alcohol and cardiovascular disease: Mendelian randomization analysis based on individual participant data. *BMJ* 2014;349: g4164.
8. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *J Am Stat Assoc* 1995;90:431–42.
9. Silverwood RJ, Holmes MV, Dale CE *et al*. Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. *Int J Epidemiol* 2014;43:1781–90.
10. Burgess S, Davies NM, Thompson SG. Instrumental variable analysis with a nonlinear exposure-outcome relationship. *Epidemiology* 2014;25:877–85.
11. Small DS. Commentary: Interpretation and sensitivity analysis for the localized average causal effect curve. *Epidemiology* 2014; 25:886–88.