# A limitation of genetic epidemiological analysis when associations are genuinely J-shaped illustrated using a prospective study of alcohol consumption and vascular disease

Professor Chris Frost‡[1]* & Professor Nicholas Wald‡[2]

Department of Medical Statistics
Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine
Keppel Street
London WCIE 7HT
UK

Institute of Health Informatics
University College London
222 Euston Road
London NW1 2DA
UK

‡ *joint first authors*

*Corresponding author: Email: chris.frost@lshtm.ac.uk

*Word count: 1207*

Keywords: Genetic Epidemiology, Alcohol, Vascular Disease, J-shaped relationship.

Mendelian randomisation, a commonly adopted approach in genetic epidemiology, is of value in distinguishing causal from non-causal associations.[1] There are, however, limitations to the conclusions that can be drawn from this approach when the underlying relationship between an outcome and an exposure is non-linear.[2] In this paper we explore a consequence of these limitations illustrating these using a prospective epidemiological study of vascular disease in 500,000 Chinese people (the Kadoorie cohort).[3] The study authors concluded that the recognised lower risk of stroke among moderate alcohol drinkers is not causal but due to confounding or reverse causation (people at risk of stroke giving up alcohol): we believe that this conclusion is potentially unsound due to the analysis methods adopted. An identical analysis of alcohol and myocardial infarction was inconclusive; a small benefit or hazard of alcohol could not be excluded.[3]

The study investigators carried out both a conventional epidemiological analysis and a genetic polymorphism analysis. The former confirmed the J-shaped dose-response relation between reported alcohol consumption and both stroke and acute myocardial infarction seen in many other studies.[4,5,6] The genetic polymorphism analysis considered two genes (ADH1B and ALDH2) that affect tolerance to alcohol, one through oxidising ethanol to acetaldehyde which causes discomfort and the other through metabolising the acetaldehyde to acetate which affects alcohol clearance; both can limit alcohol intake. Each genetic variant involves a G→A mutation and has three forms (AA, AG and GG) so the two variants yield 9 genotypes. The data were stratified according to ten geographical areas creating 90 categories that were then reduced to 6 based on the mean alcohol intake (<10, 10-, 25-, 50-, 100- and 150 or more grams per week). Individuals were assigned to one of these 6 categories (referred to as genotype-predicted alcohol intake categories) based on their genotype and study area. These categories, rather than individuals’ own reported alcohol consumption, were then related to outcomes including risk of stroke and risk of myocardial infarction.

The genetic polymorphism analysis showed no J-shaped relationship for either stroke or myocardial infarction, leading the authors to conclude that there is no safe alcohol intake threshold below which alcohol confers a health benefit. There is, however, a fundamental weakness in this analysis that tends to conceal a true underlying J-shaped relationship. The fundamental weakness arises because when a true relationship between an outcome and an exposure is non-monotonic (such as a J-shaped) then the formation of groups around the inflection point may be too coarse to reveal the non-monotonic relationship. If the groups are coarse enough the relationship may even become flat. The problem can be illustrated by examples that for simplicity assume no confounding.

Table 1 shows two simple examples based on variants in one gene and where there are just 3 categories of alcohol intake (non-drinkers, light drinkers and heavy drinkers) defined as drinking respectively 0, 100 and 500 grams of alcohol per week. In Example 1 in the table the true stroke risks over a given period of time are 15%, 10% and 20% respectively and in Example 2 are 15%, 10% and 35% respectively, so both relationships are J-shaped. There is a single gene with three variants (AA, AG, GG) that are associated with drinking as shown in the first 3 columns in section B of the table. The AA group consists entirely of non-drinkers, the AG group of an equal mixture of all three drinking categories and the GG group of an equal mixture of light and heavy drinkers. The average alcohol and stroke risks in the three genetic groups are shown in the last 2 columns of section B of the table (and in the figure for Example 1).

Though the alcohol consumption in the three drinking categories are 0, 100 and 500 grams per week respectively, averaging makes the consumptions for the three genotypes 0, 200 ($\frac{0+100+500}{3}$) and 300 ($\frac{100+500}{2}$) grams per week respectively, whilst the stroke risks are all 15% (1×15, $\frac{15+10+20}{3}$, and $\frac{10+20}{2}$) in Example 1 and 15%, 20% and 22.5% in Example 2; a monotonic relationship. The genotype analysis conceals the J-shape relationship showing no association in Example 1 and shows a continuous monotonic association in Example 2. The Figure illustrates graphically Example 1 in the Table.

Table 2 shows an example with 6 categories of the 9 variants in two genes and 7 categories of alcohol intake as was done by Millwood *et. al.* in the analysis of the Kadoorie cohort (in that paper the 6 genetic groups also took geographic area into account). As in examples 1 and 2 the genotypic categorisation converts a non-monotonic relationship into a monotonic one. Also in this example the slope of the association between alcohol consumption and stroke risk in all participants in the genotypic analysis is similar to the analogous slope in an analysis using actual alcohol consumption that is restricted to those who drink more than occasionally (illustrated by the risk ratio comparing the moderate drinking (300g/week) group with the light drinking (100g/week) group of 1.30 (13%/10%), very similar to the risk ratio comparing genotype category VI (mean consumption 315g/week) with category II (102g/week) of 1.26 (14%/11.1%)). This similarity in slopes may misleadingly appear to lend weight to the validity of the genetic epidemiological analysis.

The examples show that where an underlying association between a risk factor and the risk of a disease is not monotonic, relating mean levels of the exposure in a group not directly defined by the exposure to the mean risk of disease can systematically distort the true relationship. J- or U- shapes can be converted into monotonic relationships, and moreover this can occur with no material effect on the slope of the association above the low points of the relationship.  These effects can occur because, when a relationship between exposure and disease risk is non-linear, variation in exposure within a group (not just the mean exposure) is related to the disease risk in that group. In related work[2,7,8] other authors have explained that when exposure-outcome relationships are non-linear the estimated slopes (sometimes termed population-averaged causal effects) can only be interpreted as an average slope across the range of the exposure. What we have illustrated is that this averaging across the range of the exposure can obscure a true non-monotonic relationship.

In the illustrative examples we use here, the observation that the alcohol-disease association in the genetic epidemiological analysis is monotonic does not exclude a true J-shaped relationship. We conclude that the observation in many studies that light drinking reduces the risk of stroke but heavier drinking increases it, is not necessarily disproved by the genetic analysis of Millwood and colleagues. It is important for public health policy that the true relationship between alcohol consumption and vascular disease is recognised.

Legend to figure: Graphical representations of true and observed associations in Example 1. The top panel shows the stroke risks (as bars) and the mean alcohol consumptions (as linked circles) in three groups categorised by alcohol intake. The bottom panel shows the combinations of these alcohol intake groups that make up each of three genetic groups, and the corresponding stroke risks (as bars) and mean alcohol consumptions (as linked circles) in these genetic groups. The relationship in the top panel is J-shaped, that in the bottom panel is flat.

Conflict of Interest

None declared

**References:**

1. Davey Smith G, Hemani G. Mendelian randomisation: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23:r89-98

2. Burgess S, Davies NM, Thompson SG on behalf of EPIC-InterAct Consortium. Instrumental Variable Analysis with a Nonlinear Exposure–Outcome Relationship. *Epidemiology* 2014;25:877-885

3. Millwood IY, Walters RG, Mei XW, Guo Y, Yang L, Bian Z, *et al.* Conventional and genetic evidence on alcohol and vascular disease aetiology: A prospective study of 500 000 men and women in China. *Lancet* 2019:393:10183:1831-1842

4. Marmot MG, Shipley MJ, Rose G, Thomas BJ. Alcohol and mortality: a U-shaped curve. *Lancet* 1981;317:580-583

5. Thompson PL. J-curve revisited: cardiovascular benefits of moderate alcohol use cannot be dismissed. *MJA* 2013;198:419-422

6. Wood AM, Kaptoge S, Butterworth AS, Willeit P, Warnakula S, Bolton T, *et al.* Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599 912 current drinkers in 83 prospective studies. *Lancet* 2018;391:1513-23

7. Burgess S CHD CRP Genetics Collaboration. Identifying the odds ratio estimated by a two-stage instrumental variable analysis with a logistic regression model. *Stat Med.* 2013;32:4726-4747

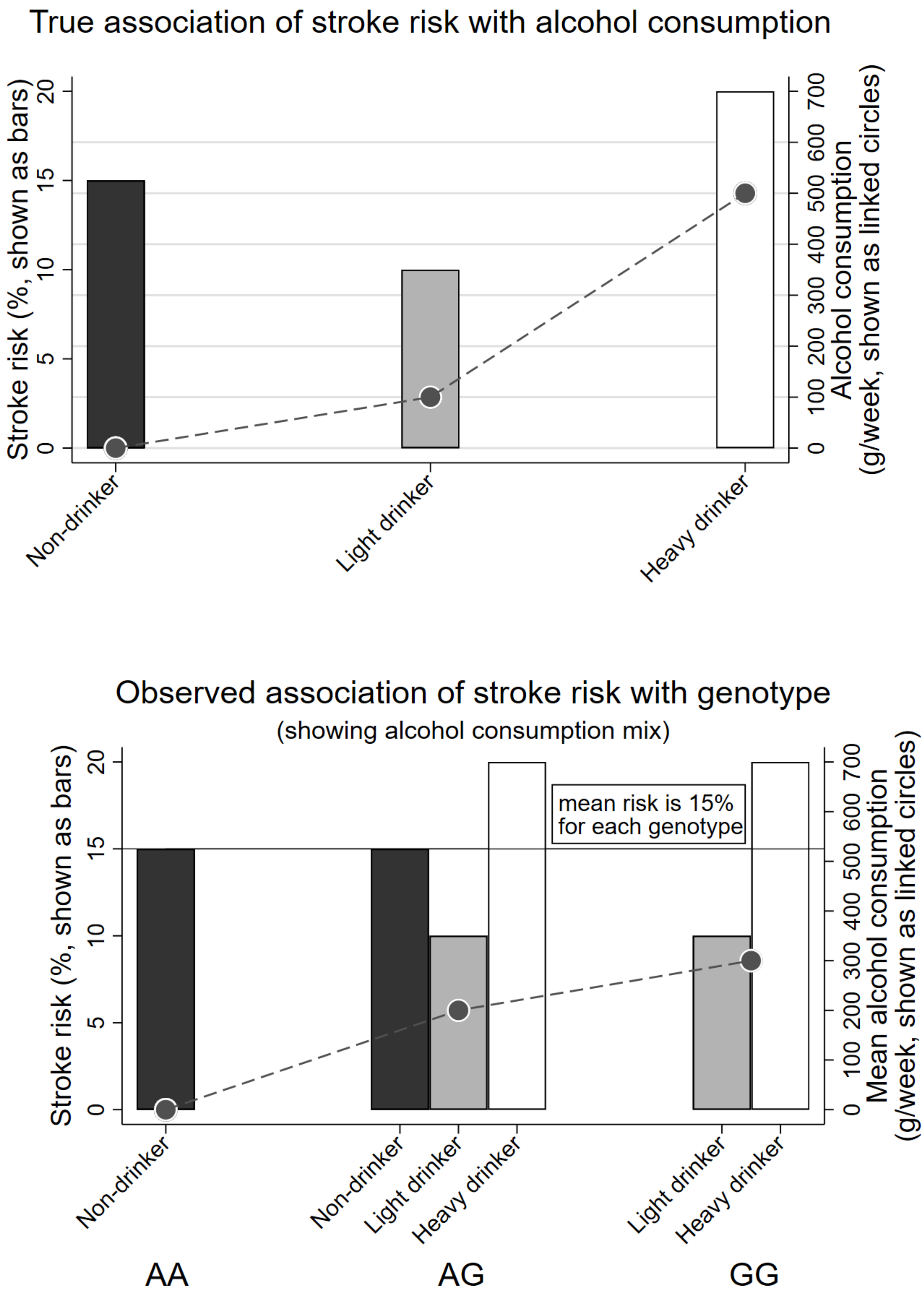
8. Mogstad M, Wiswall M. Linearity in instrumental variables estimation: problems and solutions. *Technical Report, Forschungsinstitut zur Zukunft der Arbeit.* 2010

**Table1: Alcohol and stroke associations: examples with a single gene and three categories of alcohol consumption**

| A. True association | | | | B. Genotype association | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Alcohol consumption (g/week) | Stroke risk | | Genotype | Non-drinker | Light drinker | Heavy drinker | Alcohol consumption (g/ week) | Stroke risk |
| Example 1 | | | | | | | | | |
| Non-drinker | 0 | 15% | | AA | 100% | - | - | 0 | 15% |
| Light drinker | 100 | 10% | | AG | 33.3% | 33.3% | 33.3% | 200 | 15% |
| Heavy drinker | 500 | 20% | | GG | - | 50% | 50% | 300 | 15% |
| Example 2 | | | | | | | | | |
| Non-drinker | 0 | 15% | | AA | 100% | - | - | 0 | 15% |
| Light drinker | 100 | 10% | | AG | 33.3% | 33.3% | 33.3% | 200 | 20% |
| Heavy drinker | 500 | 35% | | GG | - | 50% | 50% | 300 | 22.5% |

**Table2: Alcohol and stroke association: example 3 using six categories of variants in two genes and seven alcohol consumption categories**

| A. True association | | | | B. Genotype association | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Drinking Category | Alcohol consumption (g/week) | Stroke risk | | Genotype Category | Ex-drinker | Non-drinker | Occasional drinker | Light drinker | Modest drinker | Moderate drinker | Substantial drinker | Alcohol consumption (g/week) | Stroke risk |
| | | | | | | | | | | | | | |
| Ex- | 0 | 14% | | | | | | | | | | | |
| | | | | I | - | 30% | 30% | 40% | - | - | - | 43 | 10.6% |
| Non- | 0 | 12% | | | | | | | | | | | |
| | | | | II | - | 20% | 20% | 30% | 20% | 10% | - | 102 | 11.1% |
| Occasional | 10 | 10% | | | | | | | | | | | |
| | | | | III | - | 10% | 20% | 20% | 20% | 20% | 10% | 167 | 11.6% |
| Light | 100 | 10% | | | | | | | | | | | |
| | | | | IV | 10% | - | 10% | 20% | 20% | 20% | 20% | 211 | 12.2% |
| Modest | 200 | 12% | | | | | | | | | | | |
| | | | | V | 20% | - | - | 10% | 20% | 20% | 30% | 245 | 13.0% |
| Moderate | 300 | 13% | | | | | | | | | | | |
| | | | | VI | 30% | - | - | - | - | - | 70% | 315 | 14.0% |
| Substantial | 450 | 14% | | | | | | | | | | | |



True association of stroke risk with alcohol consumption

Observed association of stroke risk with genotype

(showing alcohol consumption mix)

Figure: