Testing for non-linear causal effects using a binary genotype in a Mendelian randomisation study: application to alcohol and cardiovascular traits

Richard J Silverwood[1,2,3], Michael V Holmes[3,4], Caroline E Dale[1], Debbie A Lawlor[5,6], John C Whittaker[1,7], George Davey Smith[5,6], David A Leon[1], Tom Palmer[8], Brendan J Keating[9], Luisa Zuccolo5[,6], Juan P Casas[1,10], Frank Dudbridge[1,2,3], on behalf of the Alcohol-*ADH1B* Consortium

1. Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK
2. Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, London, UK
3. Bloomsbury Centre for Genetic Epidemiology and Statistics, London School of Hygiene and Tropical Medicine, London, UK
4. Department of Epidemiology and Public Health, University College London, UK
5. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
6. School of Social and Community Medicine, University of Bristol, Bristol, UK
7. Genetics, R&D, GlaxoSmithKline, Stevenage, UK

8. Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK
9. Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, USA
10. Institute of Cardiovascular Science, University College London, UK

Correspondence to:

Richard J Silverwood. Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK. Phone: 0207 927 2688. Email: Richard.Silverwood@lshtm.ac.uk.

Word counts

Summary: 250

Main body: 4382

Summary

Background

Mendelian randomisation studies have so far restricted attention to linear associations relating the genetic instrument to the exposure, and the exposure to the outcome. In some cases, however, observational data suggest a non-linear association between exposure and outcome. For example, alcohol consumption is consistently reported as having a U-shaped association with cardiovascular events. In principle Mendelian randomisation could address concerns that the apparent protective effect of light-to-moderate drinking might reflect 'sick-quitters' and confounding.

Methods

The Alcohol-*ADH1B* Consortium was established to study the causal effects of alcohol consumption on cardiovascular events and biomarkers, using the single nucleotide polymorphism *rs1229984* in *ADH1B* as a genetic instrument. To assess non-linear causal effects in this study we propose a novel method based on estimating local average treatment effects for discrete levels of the exposure range, then testing for a linear trend in those effects. Our method requires an assumption that the instrument has the same effect on exposure in all individuals. We conduct simulations examining the robustness of the method to violations of this assumption, and apply the method to the Alcohol-*ADH1B* Consortium data.

Results

Our method gave a conservative test for non-linearity under realistic violations of the key assumption. We found evidence for a non-linear causal effect of alcohol intake on several cardiovascular traits.

Conclusions

We believe our method is useful for inferring departure from linearity when only a binary instrument is available. We estimated non-linear causal effects of alcohol intake which could not have been estimated through standard instrumental variable approaches.

Key words: Mendelian randomisation; Instrumental variables; Causal inference; Local average treatment effects; Alcohol consumption; Cardiovascular disease.

Key messages

- Mendelian randomisation studies have so far restricted attention to linear associations relating the genetic instrument to the exposure, and the exposure to the outcome, but this may not always be appropriate. For example, alcohol consumption is consistently reported as having a U-shaped association with cardiovascular events in observational studies.
- We propose a novel Mendelian randomisation method based on estimating local average treatment effects for discrete levels of the exposure range, then testing for a linear trend in those effects.
- Our method gave a conservative test for non-linearity under realistic violations of the key assumption in simulations, and we believe our method is useful for inferring departure from linearity when only a binary instrument is available.
- We found evidence for a non-linear causal effect of alcohol intake on several cardiovascular traits in the Alcohol-*ADH1B* Consortium, using the single nucleotide polymorphism *rs1229984* in *ADH1B* as a genetic instrument.

Introduction

Recent years have seen an increasing number of Mendelian randomisation (MR) analyses that examine causal relationships between heritable exposures, such as levels of circulating biomarkers, and outcomes such as multifactorial diseases, for example coronary heart disease and type 2 diabetes.[1,2,3] In principle MR reduces problems of confounding and abolishes reverse causation by using a genetic proxy for the exposure in an instrumental variable (IV) analysis.[4]

To date, applications of MR have been limited to linear (or log-linear) models for the associations between gene and exposure, and between exposure and outcome. In part this is because linear models have a natural interpretation which may be useful even if the true relationship is non-linear.[5] Furthermore, many of the associations between genetic variants and complex traits discovered to date have appeared to be linear.[6] However in learning about causal relationships it is clearly of value to identify and characterise non-linear effects when they are present, bearing in mind that the existence and extent of such relationships may depend on the measurement scale. In particular, non-linear associations may translate into opposing effects (protective as well as harmful) according to the level of the exposure. Such opposing effects have been observed in many observational studies examining the relationship between alcohol consumption and cardiovascular events.[7] Specifically, light-to-moderate levels of alcohol consumption have been associated with decreased risk of cardiovascular events relative to non-drinkers, with increased risk only occurring at higher levels of consumption. This apparent protective effect of light-to-moderate alcohol consumption could be explained by several different mechanisms, and corresponding 'J' or 'U' shaped associations have been observed with cardiovascular risk factors including low-density lipoprotein particles,8 abdominal adiposity,9 C-reactive protein (CRP),[10,11] and

triglycerides (TG).[12] Similar observational associations were seen in our earlier analyses of *ADH1B* Consortium data (Holmes et al, Supplementary Appendix, Figure S3).[13]

As these observational findings suggest that light-to-moderate consumption may be cardio-protective, it is of great interest to consumers, suppliers and policy makers to establish whether this pattern is causal. Confounding is plausible, since socio-economic groups that drink moderately may have other lifestyle factors that directly lead to lower rates of disease,[14] and the relationship between confounders and alcohol may themselves be non-linear. Evidence for reverse causation is also well established, with those developing ill-health or commencing medication more likely to reduce or quit alcohol consumption (the 'sick-quitters' phenomenon).[15,16]

Alcohol consumption is influenced by genetic variants that affect alcohol metabolism. Heritability of alcoholism has been estimated at 40-60%, and variants in *ALDH2*, *ADH1B* and *ADH1C* that encode for liver enzymes have been associated with decreased intake, via increased metabolism of alcohol to acetaldehyde or decreased acetaldehyde clearance, both leading to unpleasant side effects.[17] In particular, *ADH1B* has been shown to be robustly associated with alcohol consumption[18,19] and has been used in MR analyses to explore the causal effect of alcohol consumption on coronary heart disease risk factors.[20]

We recently established a large consortium (the 'Alcohol-*ADH1B* Consortium') of genetic association studies of European descent that used a single nucleotide polymorphism (SNP) in *ADH1B*, *rs1229984*, as the instrument to assess the impact of alcohol consumption on cardiovascular events and risk factors.[13] This consortium showed that carrying the *rs1229984* A-allele was associated with non-drinking, lower alcohol consumption, and lower incidence

of binge drinking, which expands the previous associations of this variant with alcohol traits.[13] Using a genetic association analysis, the consortium also showed that *ADH1B* carriers had a more favourable cardiovascular profile and a reduced risk of CHD.[13] However, because of the existing literature on non-linear effects of alcohol consumption on cardiovascular events and the lack of appropriate methods to account for non-linear associations within IV analyses, we did not initially conduct an MR analysis in the Alcohol-*ADH1B* Consortium.

Approaches have been proposed for non-linear IV analysis in the econometric literature,[21,22,23] but they cannot be used in this context because we use a single SNP as the IV. In the present paper, we develop new methods to conduct non-linear IV analysis using a single binary instrument, and also evaluate the impact of the key assumption of our method. We then apply our method to the data from the Alcohol-*ADH1B* Consortium to assess whether the causal effect of alcohol on cardiovascular traits is indeed non-linear and whether this implies a non-zero optimal level of consumption for cardiovascular health, which has clear implications for public health.

Material and methods

Data

The Alcohol-*ADH1B* Consortium is a collaboration of studies in which the associations between an allele of the *ADH1B* gene and twenty-two cardiovascular biomarkers, risk of coronary heart disease, stroke and type 2 diabetes have been examined.[13] Here our analyses are restricted to the 22 studies (18 cohorts, 2 nested case-control studies, 1 randomized trial and 1 case-control study) with individual participant data originating from Europe (n = 16) and North America (n = 6). Analysis was restricted to individuals of European descent.[13]

The principal alcohol trait was weekly volume of alcohol in British units (1 British unit is equivalent to 0.57 US units or 10 ml [7.9g] ethanol), which we derived using questionnaire data from each study. For studies in which this variable was not already present, we either calculated weekly volume of alcohol by summing over the individual components of beverage-specific drink questions (available in 20 of the 22 studies), or by converting alcohol recorded in grams/week into British units.[13] The units/week were log transformed, after incrementing by one to allow for individuals reporting zero weekly alcohol consumption, resulting in a normally distributed phenotype that had homoscedastic residual error after regressing on the *ADH1B* genotype.

Here we considered a subset of outcomes for which a non-linear causal association was either postulated from subject-matter knowledge, or suggested by the observational data available from the Alcohol-*ADH1B* Consortium (all $P < 0.001$ for the quadratic term in a quadratic model): systolic blood pressure (SBP), non-high density lipoprotein cholesterol (non-HDL-C), TG, high density lipoprotein cholesterol (HDL-C), body mass index (BMI), waist circumference (WC), CRP and interleukin 6 (IL-6). Outcomes were log transformed towards normality when appropriate (TG, CRP and IL-6).

The *rs1229984* polymorphism in *ADH1B* was directly genotyped in all studies and coded as 0/1 according to the carriage of at least one minor allele. This coding was adopted owing both to the low prevalence of the *rs1229984* A-allele (average carriage of *rs1229984* A-alleles in the analysis sample: 7.7%) and the stronger association observed with alcohol dependence and other alcohol-related traits under a dominant model compared to a recessive model.[24]

Full details of participating studies, phenotype definition and genotyping are reported elsewhere[13] and are summarised in Table S1 in the Supplementary Data.

Linear instrumental variable analysis

We used standard two-stage least squares (2SLS) to estimate a linear causal effect of log(weekly units of alcohol + 1) (hereafter, log-alcohol) on continuous cardiovascular outcomes. That is, we fitted the first stage linear regression

$$x_i = \beta_{XG} g_i + \beta_{XZ}^{'} z_i + \varepsilon_{Xi}$$

where x_i is log-alcohol for subject i, g_i is a binary code for the *rs1229984* genotype, z_i is a vector of covariates and ε_{Xi} are residual errors assumed to be independent and identically distributed with mean zero. Regression coefficients β_{XG} and β_{XZ} were estimated as fixed effects. We used the fitted model to predict $\hat{x}_i$ then estimated the alcohol-outcome association β_{YX} from the regression

$$y_i = \beta_{YX} \hat{x}_i + \beta_{YZ}^{'} z_i + \varepsilon_{Yi}$$

where y_i is the continuous cardiovascular outcome for subject i and ε_{Yi} are residual errors assumed to be independent and identically distributed with mean zero. A 95% confidence interval for $\hat{\beta}_{YX}$ was derived by nesting the 2SLS within a bootstrap resampling procedure

using 10 000 bootstrap samples. As covariates we included in both regressions a fixed effect for each study, and fixed effects for age and sex.

Non-linear causal effects

To test for non-linearity of the causal *X*-*Y* association we consider local average treatment effects (LATEs) in subgroups of *X*.[25] First we coarsen *X* into a discrete and rescaled variable $X^* = \left\lfloor \frac{X}{\beta_{XG}} \right\rfloor$ with finite support, assumed without loss of generality to be $\{0,\ldots,J\}$ for fixed *J*. *G* is an instrument for X^* if it is independent of the remainder $X - X^*$ (see Figure 1); this is not generally true but it can be tested in applications. Under linear models we can obtain an estimate of the causal effect of X^* on Y, but this effect can also be represented as a weighted sum of LATEs,[25,26] which are causal effects among the individuals whose exposures X^* are changed from one level to the next by the genetic instrument.

More precisely, let $Y_i(j)$ denote the potential outcome for subject *i* obtained by setting, possibly contrary to fact, the exposure $X_i^* = j$. Moreover let $X_i^*(0)$ and $X_i^*(1)$ be the possibly counterfactual values of the exposure obtained by setting the binary instrument to 0 and 1 respectively. Then the LATE at exposure level *j* is defined as

$$\tau_j = E[Y_i(j) - Y_i(j-1) \mid X_i^*(1) \geq j > X_i^*(0)]$$

that is, the average treatment effect among those whose exposure would be at least *j* if their instrument were set to 1, and whose exposure would be less than *j* if their instrument were set to 0. Identification of LATEs requires the further assumption of monotonicity, that is either

$X_i^*(1) - X_i^*(0) \geq 0$ or $X_i^*(1) - X_i^*(0) \leq 0$ for all subjects *i*, implying that the instrument either does not decrease the exposure in all subjects, or does not increase it in all subjects.

[Figure 1 here]

If we could estimate the LATEs τ_j then testing them for equality would provide a direct test of linearity of the causal effect. Here we propose an assumption that allows this to be performed. Assume that the causal effect of the instrument on the discretised exposure is exactly 1 in each subject:

$$X_i^*(1) - X_i^*(0) = 1 \forall i$$

This is a stronger version of the monotonicity assumption. In fact, this assumption will hold if the first stage linear model is a true structural model for *X*, with no unmeasured confounders of the *G*-*X* association, or modifiers of the effect of *G* on *X*. Under this assumption (and noting that X has been rescaled so that a one unit change in X^* corresponds to the expected exposure change with genotype), every subject contributes to a LATE, since for every *i* there is a *j* such that $X_i^*(1) \geq j > X_i^*(0)$, in fact $X_i^*(1) = j = X_i^*(0) + 1$. That is, the instrument moves each subject from one level of X^* to the next: in the randomised trials terminology, all subjects are compliers.

It is now possible to assign each subject to the estimation of a LATE, based on the observed data. Since $X_i^*(1) = j = X_i^*(0) + 1$ if and only if $X_i^* = j$ and $G_i = 1$ or $X_i^* = j - 1$ and $G_i = 0$, we can write the LATE as

$$\begin{aligned}\tau_j &= E[Y_i(j) - Y_i(j-1) \mid X_i^*(1) \geq j > X_i^*(0)] \\ &= E[Y_i(j) \mid X_i^* = j, G_i = 1 \vee X_i^* = j-1, G_i = 0] - E[Y_i(j-1) \mid X_i^* = j, G_i = 1 \vee X_i^* = j-1, G_i = 0] \\ &= E[Y \mid X_i^* = j, G_i = 1] - E[Y \mid X_i^* = j-1, G_i = 0]\end{aligned}$$

which may be estimated using ordinary linear regression (possibly with adjustment for relevant covariates) restricted to the subjects having $X_i^* = j$ and $G_i = 1$ or $X_i^* = j-1$ and $G_i = 0$.

Having estimated a LATE (with its standard error) for each level of X^*, the estimates may be tested for equality using standard methods of meta-analysis. In particular, we use meta-regression to test for a linear trend in the LATEs. A linear model relating LATEs to the exposure levels

$$E(\tau_j) = \gamma_1 + \gamma_2 j$$

would apply if the underlying causal model were quadratic

$$E(Y) = \gamma_0 + \gamma_1 j + \tfrac{1}{2}\gamma_2 j^2$$

The coefficient γ_2 is zero if the LATEs are equal, which is the case when the causal effect of *X* on *Y* is linear. Then the mean LATE, calculated by fixed-effects meta-analysis of the estimated LATEs, is an alternative measure of the linear causal effect of *X*. Rejection of $\gamma_2 = 0$ implies a non-linear causal effect; a quadratic form is not directly implied but such a model could be hypothesised, up to its intercept term, from the fitted meta-regression. The

estimation of a linear model relating LATEs to the exposure levels is a simple but powerful way to investigate departures from linearity, as any such departures are captured by a single parameter. However, alternative models could be fitted to characterise the dose-response relationship more flexibly. For example, a piecewise constant model relating the LATEs to the exposure levels would correspond to a linear spline model relating the exposure to the outcome. This could be detected by a test of Cochran's Q on the estimated LATEs.

This procedure requires rescaling of *X* by the effect size β_{XG} of the instrument. However the true value of β_{XG} is unknown and it must be estimated. To account for sampling uncertainty in $\hat{\beta}_{XG}$ we nest the entire LATE and meta-regression procedure within a bootstrap resampling procedure, using 10 000 bootstrap samples, to obtain proper confidence intervals on the meta-regression estimates $\hat{\gamma}_1, \hat{\gamma}_2$. Our procedure for testing departure from linearity of the causal effect of *X* on *Y* is summarised in Box 1.

[Box 1 here]

Beyond a test for departure from linearity, we are interested in identifying the way the causal effect changes with increasing alcohol consumption and, in particular, the nadir of the curve which could be conceived as an 'optimal' level of consumption regarding cardiovascular traits. As we cannot estimate the intercept term in the fitted quadratic model, we cannot predict the absolute value of the outcome for a given level of alcohol consumption, so we focus on the difference in outcome relative to zero alcohol consumption. For those outcomes with evidence of non-linearity we predict this at four values of alcohol consumption (3.04, 12.15, 31.90 and 84.52 units/week), which are the medians of observed values in the categories representing low (>0-7 units/week), moderate (7-21 units/week), heavy (21-70

units/week) and very heavy (70+ units/week) alcohol consumption in the analysis of Holmes et al.[9] By differentiation of the hypothesised quadratic function we estimate three additional features of the curve: i) the 'optimal' level of alcohol consumption; ii) the difference in outcome at the optimal alcohol consumption relative to zero alcohol consumption; iii) the level of alcohol consumption required to have an outcome level equivalent to that at zero alcohol consumption. Confidence intervals for all the estimates are obtained by nesting the estimation within the bootstrap resampling procedure outlined above. In the bootstrap samples we left truncated the nadir of alcohol consumption at zero.

All analyses were conducted using R version 2.13.[27]

Simulations

We conducted simulations to assess the proposed approach in terms of bias and coverage under various data generating models. Full details and results are given in the Supplementary Data. In brief, we simulated data in which there was no causal *X*-*Y* association, in which the association was linear, and in which there was a quadratic causal association, allowing throughout for quadratic effects of confounders. We assessed robustness to the assumption of individual level homogeneity of the genetic effect using additional simulations of β_{XG} heterogeneity and *G*-*U* interaction at both the individual and subgroup level.

We observed that the LATE estimates were essentially unbiased with generally good coverage properties under null, linear and quadratic models, and that the test for a non-linear effect was slightly conservative. Together the results suggest that this method is a useful extension to standard approaches in the non-linear setting. Reasonable levels of individual-

level heterogeneity in β_{XG} or between-subgroup heterogeneity in β_{XG} were not found to lead to significant bias in the estimates. High levels of interaction between G and U led to bias in the estimates, but such interactions may be unlikely in practice.

Results

We investigated the potential non-linear effects of log-alcohol on each of the outcomes in the Alcohol-*ADH1B* Consortium using the proposed procedure. Some issues relating to the inclusion of multiple studies in the Consortium are discussed in the Supplementary Data.

Age- and sex-adjusted study-specific estimates of the association between *rs1229984* and log-alcohol are presented in Figure S17 of the Supplementary Data. These study-specific estimates have (inverse-variance-weighted) mean -0.235 and SD 0.121, indicating some degree of between-study variability. However, in our simulations (see Supplementary Data) a similar degree of heterogeneity between known subgroups (scenario 'f' with $\gamma = 0.1$) was not found to result in bias to either the LATE intercept or slope, with slightly conservative confidence intervals for each.

To examine whether *G* = *rs1229984* is a valid instrument for discretised X^*, assuming that it is valid for the continuous measure X = log-alcohol, we examined the correlation between *G* and the remainder $X - X^*$; these should be independent for *G* to be a valid instrument for X^*. We observed a weak but significant correlation (Pearson's *r* = -0.013, 95% CI: -0.020, -0.006). We hypothesised that this residual correlation was due to the large number of individuals reporting drinking zero weekly units of alcohol (log-alcohol = 0), because these

individuals have a residual $X - X^* = 0$ and are also more likely to have $G = 0$. When individuals with log-alcohol = 0 were excluded from the analysis the correlation between G and the remainder $X - X^*$ was close to zero (Pearson's r = 0.001, 95% CI: -0.007, 0.009). We therefore re-analysed the data after excluding individuals with log-alcohol = 0, but obtained very similar results to those from the full sample. Because it is necessary to retain individuals reporting zero drinking to meet the objectives of the analysis, we only report results using the full sample.

The results of the LATE-based analysis for each of the outcomes are presented in Table 1 along with the standard linear IV analysis. We illustrate our approach in more detail using SBP as an example, following the steps in Box 1. We estimated $\hat{\beta}_{XG} = -0.244$ assuming a common genetic effect across all studies. Discretising log-alcohol into units of -0.244 gave an integer exposure X^* with range [-26, 0]. We then estimated the LATE at each value of X^*. For example, for $j = -11$ (corresponding to a log-alcohol of -11 × -0.244 = 2.684, or exp(2.684) – 1 = 13.6 units/week) we selected the subjects with $X^* = -11$ and *rs1229984* = 1, or $X^* = -12$ and *rs1229984* = 0. Linear regression of SBP on X^*, on these subjects only, and adjusting for study, age and sex, gave $\tau_{-11} = -1.55$. That is, in subjects whose X^* was changed from -12 to -11 by the SNP, their SBP was decreased by 1.55 mmHg.

[Table 1 here]

Rescaling by $\hat{\beta}_{XG} = -0.244$, subjects whose log-alcohol was changed from $-12 \times -0.244 = 2.928$ to $-11 \times -0.244 = 2.684$ (i.e. whose weekly units of alcohol consumption was changed from exp(2.928) – 1 = 17.7 to exp(2.684) – 1 = 13.6) by the SNP

have their SBP decreased by 1.55 mmHg. Alternatively, a one unit *increase* in log-alcohol at this level of alcohol consumption (e.g. from 2.684 to 3.684, or from exp(2.684) – 1 = 13.6 to exp(3.684) – 1 = 38.8 units/week – a considerable increase) was associated with an increase in SBP of $-1.55/-0.244 = 6.35$ mmHg.

The full graph of estimated LATEs for SBP is shown in Figure 2. Negative LATEs represent decreasing SBP with log-alcohol whilst positive LATEs represent increasing SBP, so a LATE trend crossing zero from negative to positive indicates a nadir. Fixed effects meta-analysis of these effects gave a mean LATE of 4.9 (95% CI: 2.6, 7.5), which is effectively a complier average treatment effect and similar to the linear IV estimate of 5.2 (95% CI: 3.2, 7.3). Meta-regression of the estimated LATEs on X^* gave a slope of 3.3 (95% CI: 1.0, 5.5). This provided strong evidence (Z-test $p = 0.004$) that the LATEs were not constant across values of log-alcohol; that is, there was a non-linear association between log-alcohol and SBP.

[Figure 2 here]

Full results for the remaining outcomes are provided in Table 1. As indicated by the LATE slope there was evidence of a non-linear causal effect for SBP, non-HDL-C, BMI, WC and CRP (all $P \leq 0.01$). For other outcomes there was no evidence of a non-linear causal effect (HDL-C, IL-6 and triglycerides, all $P > 0.4$, though note that power is lower for IL-6 due to the relatively smaller sample size). In these cases we recommend that the linear IV results are employed as fewer assumptions are required in their estimation. It should also be noted that the linear IV estimates and the mean LATEs were similar for each of the outcomes, albeit with the latter having wider CIs.

Table 2 shows the predicted difference in each outcome relative to zero alcohol consumption for 3.04, 12.15, 31.90 and 84.52 units/week of alcohol consumption under the fitted quadratic functions. All outcomes, with the exception of SBP, were predicted to be lower at 3.04 units/week ('low' alcohol consumption) than at zero alcohol consumption, though each confidence interval included the possibility of no true difference. By 31.90 units/week ('heavy' alcohol consumption) all outcomes were predicted to be higher than at zero alcohol consumption, though each confidence interval, with the exception of SBP, again included the possibility of no true difference. By 84.52 units/week ('very heavy' alcohol consumption) all the confidence intervals excluded the possibility of no true difference.

[Table 2 here]

Table 2 also shows the additional estimated features of the hypothesised quadratic functions. For all outcomes the optimal level of alcohol consumption was estimated to be greater than zero, ranging from 1.0 units/week (SBP) to 3.5 units/week (CRP). However, only for non-HDL-C did the confidence interval exclude the possibility that zero consumption may be optimal. Correspondingly, the estimated difference in outcome at the optimal alcohol consumption level relative to zero consumption was negative for each outcome, though only for non-HDL-C did the confidence interval exclude the possibility of no true difference. The level of alcohol consumption required to have an outcome level equivalent to that at zero consumption was estimated as ranging from 2.8 units/week (SBP) to 19.4 units/week (CRP), though for all outcomes the confidence intervals were very wide. These results are illustrated for non-HDL-C, for which the strongest evidence of non-linearity was observed, in Figure 3. However, the precise values of our quantitative results should be interpreted with some

caution as the quadratic causal model that we fit may not be sufficiently flexible to fully characterise the dose-response relationship.

[Figure 3 here]

Discussion

We have proposed a method based on estimating LATEs that allows a basic estimation of local causal effects of a continuous exposure when using a binary instrument. Our method requires an assumption of homogeneous individual treatment effects of the instrument on the exposure, but our simulations found the estimates obtained under our approach to be largely unbiased and with good coverage properties under a variety of heterogeneous effects of instrument on exposure.

The local effects we estimate are within discretised units of the exposure, with the size of those units depending on the gene-exposure association. This is not a scale with a generally useful interpretation, and different genetic instruments could lead to different discrete units with different definitions of local causal effects. We therefore emphasise the ability to test for a non-linear causal effect and draw qualitative conclusions about the shape of that effect, and we suggest that a strictly quantitative interpretation of the estimated parameters should be viewed with some caution. Further work is required in investigating alternative models relating the LATEs to the exposure levels in order to provide greater flexibility for characterising the dose-response relationship.

Using this approach we detected evidence for a non-linear causal effect of log-alcohol on several cardiovascular traits in a large collaborative study, which would not have been possible using standard IV approaches. For each outcome that exhibited evidence of a non-linear causal effect, our results suggested that the level of alcohol consumption associated with the lowest value of the cardiovascular traits to lie between 1.0 and 3.5 units/week. However, only for non-HDL-C do we have strong evidence that the optimal level of consumption truly differs from zero.

As the cardiovascular traits considered in this analysis were observed concurrently with the level of alcohol consumption in many of the studies within the *ADH1B* Consortium, a conventional analysis would be at risk of bias due to reverse causality (for example, someone with high SBP reducing their alcohol intake so that they are observed to have a low level of consumption). A Mendelian randomisation analysis removes the possibility of such reverse causality, which is a significant strength of the present study.

For our estimated effects to be interpreted causally we need the standard assumptions underlying MR analysis to hold. Of particular concern in the present application is the exclusion restriction that *G* has no effect on *Y* other than through *X*. We have only considered one aspect of alcohol consumption (weekly units), but if the polymorphism in *ADH1B* reduces alcohol consumption generally then other aspects, such as frequency of binge drinking, may also be associated with the instrument.[19] If such other aspects have a causal effect on the outcome independently of weekly units then the exclusion restriction would not hold. The strong correlation between weekly units and other aspects of alcohol consumption makes a significant violation of this assumption unlikely. However, further research is required in this area.

Although we limited our analyses to individuals of European descent and adjusted for study in all our analyses, there may be residual population stratification of the variant which could lead to backdoor pathways from the instrument to the outcome. The restriction to individuals of European descent may also reduce the generalisability of our findings beyond such populations.

An inherent aspect of our approach is the need for a large sample with a sufficiently strong association between the gene and the exposure. If the gene-exposure association is very weak then the exposure will be discretised into many bins, none of which will contain sufficient subjects for the LATEs to be estimated. Many MR studies are now conducted on large samples in order to improve power to detect causal effects, but our approach requires large samples across a sufficient range of the exposure in order to detect non-linearities. This problem is compounded when studying binary outcomes, as each bin should contain a sufficient number of events. Therefore we have restricted our attention to continuous outcomes in this paper, but we recognise that here the key interest is in the nature of the causal relationship with cardiovascular disease events, which cannot be readily deduced from the associations with different risk factors. Further work in this area is required.

We believe our method is useful for inferring departure from linearity when only a binary instrument is available. Although there is clearly greater scope for bias than in standard IV analysis, we did not infer non-linear effects for several of the cardiovascular outcomes we considered, suggesting some degree of specificity using our method. More robust inference of non-linear causal effects may be possible from polychotomous or continuous instruments, such as gene scores constructed from multiple SNPs.[28,29] Such instruments will allow the identification of non-linear models with many parameters, though IV estimation of parametric non-linear models has been found to be dependent on the choice of parametric model.[23] A further

key issue is whether the exposures predicted by those instruments cover a sufficient range to capture the non-linear features of the causal effects. If this is not the case, then it may be necessary to pursue approaches based on local effects similar to the one for binary instruments that we have discussed here.

Acknowledgements

This work was supported by the UK Economic and Social Research Council [NCRM Pathways node ES/I025561/2 to R.S.] and the UK Medical Research Council [Population Health Scientist Fellowship G0802432 to M.V.H and G1000718 to F.D.]. D.A.L. and G.D.S. work in a Unit that receives funding from the UK Medical Research Council [MC_UU_12013/1-9] and the University of Bristol.

Alcohol-*ADH1B* Consortium Acknowledgements

ARIC: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number

UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research; **BRHS**: The British Regional Heart Study has been supported by programme grant funding from the British Heart Foundation (RG/08/013/25942); **BWHHS**: The British Women's Heart and Health Study has been supported by funding from the British Heart Foundation (BHF) (grant PG/09/022) and the UK Department of Health Policy Research Programme (England) (grant 0090049). The BWHHS HumanCVD data were funded by the BHF (PG/07/131/24254).; We thank all BWHHS participants, the general practitioners and their staff who have supported data collection since the study inception; **CARDIA** is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging; **CHS**: This research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01 HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, N01HC65226, and grant HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org; **Cyprus**: The Cyprus Study has been supported by the Cyprus Cardiovascular Disease Educational and Research Trust (CCDERT) and Joint Cyprus Research Promotion Foundation, Ministry of Health and Cyprus Heart Foundation grant No 41/5PE as well as Research Promotion Foundation grants (PENEK 05/04 and YGEIA 04/06); **Czech post-**

MONICA: Supported by MH CZ - DRO ("Institute for Clinical and Experimental Medicine - IKEM, IN 00023001"). **EAS**: The EAS was funded by the British Heart Foundation (Programme Grant RG/98002); **ELSA**: Samples from the English Longitudinal Study of Ageing (ELSA) DNA Repository (EDNAR), received support under a grant (AG1764406S1) awarded by the National Institute on Ageing (NIA). ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute of Fiscal Studies. The data were collected by the National Centre for Social Research.; **EPIC Turin**: The EPIC Turin study is funded by grants from the Associazione Italiana per le Ricerche sul Cancro, Italy and grants from the Compagnia di San Paolo, Turin, Italy; **FHS**: The Framingham Heart Study began in 1948 with the recruitment of an original cohort of 5,209 men and women (mean age 44 years; 55 percent women). In 1971 a second generation of study participants was enrolled; this cohort consisted of 5,124 children and spouses of children of the original cohort. The mean age of the offspring cohort was 37 years; 52 percent were women. A third generation cohort of 4,095 children of offspring cohort participants (mean age 40 years; 53 percent women) was enrolled beginning in 2002. At each clinic visit, a medical history was obtained with a focus on cardiovascular content, and participants underwent a physical examination including measurement of height and weight from which BMI was calculated; **HAPIEE**: This study was supported by Wellcome Trust 'Determinants of Cardiovascular Diseases in Eastern Europe: A multi-centre cohort study' [grants 064947/Z/01/Z; and 081081/Z/06/Z]; the MacArthur Foundation 'MacArthur Initiative on Social Upheaval and Health' [grant 712058]; the National Institute on Ageing 'Health disparities and aging in societies in transition (the HAPIEE study)' [grant 1R01 AG23522]; and a project from the Ministry of Health, Czech Republic, for the development of the research organization No. 00023001 (IKEM, Prague, Czech Republic). We would like to thank researchers, interviewers and participants in Novosibirsk, Krakow, Kaunas,

Havířov/Karviná, Jihlava, Ústí nad Labem, Liberec, Hradec Králové, and Kroměříz.; **Inter99**: The Inter99 study was supported by the Danish Medical Research Council, the Danish Centre for Evaluation and Health Technology Assessment, Copenhagen County, the Danish Heart Foundation, the Danish Pharmaceutical Association, the Health Insurance Foundation, the Augustinus Foundation, the Ib Henriksens foundation and the Beckett Foundation. The present study was further supported by the Danish Diabetes Association (grant No. 32, December 2005) and the Health Insurance Foundation (grant No. 2010 B 131); **Izhevsk**: The Izhevsk Family Studies was funded by a UK Wellcome Trust programme grant (078557); **MESA**: The Multi-Ethnic Study of Atherosclerosis Study (MESA) is a multicenter prospective cohort study initiated to study the development of subclinical cardiovascular disease. A total of 6814 women and men between the age of 45 and 84 year were recruited for the first examination between 2000 and 2002. Participants were recruited in six US cities (Baltimore, MD; Chicago, IL; Forsyth County, NC; Los Angeles County, CA; Northern Manhattan, NY; and St. Paul, MN). This study was approved by the institutional review boards of each study site, and written informed consent was obtained from all participants. This cohort was genotyped as part of the National Heart Lung and Blood Institute's (NHLBI) Candidate Gene Association Resource (CARe) (Musunuru, K., Lettre, G., Young, T., Farlow, D.N., Pirruccello, J.P., Ejebe, K.G., Keating, B.J., Yang, Q., Chen, M.H., Lapchyk, N. et al. Candidate gene association resource (CARe): design, methods, and proof of concept. Circ. Cardiovasc. Genet, 3, 267-275.); **NPHS II**: NPHS-II was supported by the British Medical Research Council, the US National Institutes of Health (grant NHLBI 33014), and Du Pont Pharma, Wilmington, Delaware; **Whitehall II**: The Whitehall II study and Mika Kivimaki were supported by the Medical Research Council; the British Heart Foundation; the Economic and Social Research Council; the National Heart Lung and Blood Institute (NHLBI: HL36310); and the National Institute on Aging (AG13196), US, NIH; **WHI**: The

WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. A listing of WHI investigators can be found at https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

Alcohol-*ADH1B* Consortium

Michael V Holmes1, MRCP, MSc, Caroline E Dale2, PhD, Luisa Zuccolo3,4, PhD, Richard J Silverwood2, PhD, Yiran Guo5,6, PhD, Zheng Ye7, PhD, David Prieto-Merino2, PhD, Abbas Dehghan8, MD, PhD, Stella Trompet9, PhD, Andrew Wong[10], Alana Cavadino[11], MSc, Kieran McCaul[12], PhD, Dagmar Drogan[13], MPH, Sandosh Padmanabhan[14], MBBS, PhD, Shanshan Li[15], MD, MSc, Ajay Yesupriya[16], MPH, Maarten Leusink[17], MSc, Johan Sundstrom[18], MD, PhD, Hynek Pikhart1, PhD, Daniel I Swerdlow1, PhD, Andrie G Panayiotou[19], PhD, Svetlana A. Borinskaya[20], PhD, Chris Finnan1, PhD, Sonia Shah[21], MSc, Karoline B Kuchenbaecker[22], Dipl.Pysch., MSc, Tina Shah1, PhD, Jorgen Engemann1, Lasse Folkerson[23], PhD, Per Eriksson[23], PhD, Fulvio Ricceri[24], PhD, Olle Melander[25], MD, PhD, Carlotta Sacerdote[26], MD, PhD, Dale M Gamble[27], MHSc, CCRP, Sruti Rayaprolu[28], BSc, Owen A Ross[28], PhD, Stela McLachlan[29], PhD, Olga Vikhireva1, MD, PhD, Ivonne Sluijs[30], PhD, Robert A Scott7, PhD, Vera Adamkova[31], MD, Leon Flicker[12], FRACP, PhD, Frank M van Bockxmeer[32], FAHA, Pedro Marques-Vidal[33], MD, PhD, Tom Meade2, FRS, Sir Michael G Marmot[34], FRCP, PhD, Jose M Ferro[35,36], MD, Sofia Paulos-Pinheiro[37,38], Steve

Humphries[39], FRCPath, FRCP, PhD, Philippa Talmud[39], FRCPath, DSc, Irene Mateo Leach[40], PhD, Niek Verweij[40], MSc, Allan Linneberg[41], MD, PhD, Tea Skaaby[41], MD, Pieter A Doevendans[42], MD, PhD, Maarten J Cramer[42], MD, PhD, Pim van der Harst[40, 43, 44], MD, PhD, Olaf H Klungel[17], PharmD, PhD, Nicole F Dowling[16], PhD,Anna F Dominiczak[14], OBE, FRCP, FRSE, FAHA, Meena Kumari1, PhD, Andrew Nicolaides[45,46,47], FRCS, PhD, Shah Ebrahim2, [48], FRCP, DM, Tom R Gaunt4, PhD, Jackie F Price[29], MD, Lars Lannfelt[49], MD, PhD, Anne Peasey1, PhD, Anke H Maitland-van der Zee[17], PhD, Paul E Norman[50], MD, Graeme J Hankey[51, 52], MD, FRACP, Manuela M Bergmann[13], PhD, Albert Hofman8, MD, PhD, Oscar H Franco8, MD, PhD, Jackie Cooper[53], MSc, Jutta Palmen[39], PhD, Wilko Spiering[54], MD, PhD, Pim de Jong[55], MD, PhD, Diana Kuh[10], PhD, Rebecca Hardy[10], PhD, Andre G UItterlinden8, PhD, Arfan M Ikram8, MD, PhD, Ian Ford[56], PhD, Elina Hypponen[11, 57], PhD, Osvaldo Almeida[12, 58, 59], MD, Nicholas J Wareham7, MB, PhD, Kay-Tee Khaw[60], FRCP, PhD, Anders Hamsten[23, 61], FRCP, PhD, on behalf of IMPROVE study group[62], Lise Lotte N Husemoen[41], PhD, Anne Tjønneland[63], MD, PhD, Janne S Tolstrup[64], PhD, Eric Rimm[15], Sc.D., Jaroslav Hubacek[31], DSc, PhD, Joline WJ Beulens[30], PhD, WM Monique Verschuren[65], PhD, N Charlotte Onland-Moret[30, 66], PhD, Marten H Hofker[67], PhD, S. Goya Wannamethee[68], PhD, Peter H Whincup[69], PhD, FRCP, Richard Morris[68], PhD, Astrid M Vicente[37, 70, 68], PhD, Hugh Watkins[72, 73], FRCP, PhD, Martin Farrell[72, 73], FRCPath, J Wouter Jukema9, [42, 44], MD, PhD, James Meschia[28], MD, L Adrienne Cupples[74, 75], PhD, Stephen J Sharp7, Msc, The InterAct Consortium[76], Myriam Fornage[77], PhD, Matthew B Lanktree[78], BSc, David S. Siscovick[79], MD, MPH, Eric Jorgenson[80], PhD, Bonnie Spring[81], PhD, Josef Coresh[82], MD, PhD, MHS, Yun R Li5, BSc, Sarah G Buxbaum[83], PhD, Pamela J Schreiner[84], PhD, R Curtis Ellison[85], MD, Michael Y Tsai[86], MD, PhD, Sanjay R Patel[93], MD, Susan Redline[15], MD, MPH, Andrew D Johnson[75], PhD, Ron C Hoogeveen[87], PhD, Hakon Hakonarson4, MD, PhD, Jerome I. Rotter[88], MD, Eric Boerwinkle[89], PhD, Paul IW de

Bakker[30, 90], PhD, Mika Kivimaki1, PhD, Folkert Asselbergs[30, 42, 90], MD, PhD, Naveed Sattar[91], FRCP, Debbie Lawlor3,[4], PhD, John Whittaker2,[92], PhD, George Davey Smith3,[4], MD, DSc, Ken Mukamal[93], MD, MPH, Bruce Psaty[79, 94], MD, PhD, MPH, James G Wilson[95], MD, PhD, Leslie A Lange[96], PhD, Ajna Hamidovic[97], PharmD, Aroon D Hingorani1, FRCP, PhD, Børge G Nordestgaard[98, 99, 100], MD, DMSc, Martin Bobak1, MD, PhD, David Leon2, PhD, Claudia Langenberg7, MD, PhD, Tom Palmer[101], PhD, Alex P Reiner[102], MD, MSc, Brendan J Keating5, PhD, Frank Dudbridge2, PhD, Juan P Casas1,[2], MD, PhD

1. Genetic Epidemiology Group, Institute of Cardiovascular Science, Department of Epidemiology and Public Health, University College London, UK.
2. Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK.
3. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
4. School of Social and Community Medicine, University of Bristol, Bristol, UK.
5. Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, PA, USA.
6. BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China.
7. MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK.
8. Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.
9. Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands.
10. MRC National Survey of Health and Development, MRC Unit for Lifelong Health and Ageing at UCL, London, UK.
11. Centre for Paediatric Epidemiology and Biostatistics, MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, London, UK.

12. Western Australian Centre for Health & Ageing, Centre for Medical Research, University of Western Australia, Perth, Western Australia, Australia.
13. German Institute of Human Nutrition Potsdam-Rehbrücke, Nuthetal, Germany.
14. Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom.
15. Department of Epidemiology, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA.
16. Office of Public Health Genomics, Office of Epidemiology, Surveillance, and Laboratory Services, Centers for Disease Control and Prevention, Atlanta, GA, USA.
17. Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands.
18. Department of Medical Sciences, Uppsala University, Uppsala University Hospital, Uppsala, Sweden.
19. Cyprus International institute for Environmental and Public Health in association with the Harvard School of Public Health, Cyprus University of Technology, Limassol, Cyprus.
20. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia.
21. UCL Genetics Institute, Department of Genetics Environment and Evolution, London, UK.
22. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
23. Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.
24. HuGeF Foundation, Torino, Italy.
25. Department of Clinical Sciences, Lund University, Malmö, Sweden.

26. Unit of Cancer Epidemiology, San Giovanni Battista Hospital and Center for Cancer Prevention (CPO-Piemonte), Torino, Italy.
27. Mayo Clinic, Jacksonville, FL, USA.
28. Department of Neuroscience, Mayo Clinic Florida, Jacksonville, FL, USA.
29. Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK.
30. Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands.
31. Center for Experimental Medicine, Institute for Clinical and Experimental Medicine, Prague, Czech Republic.
32. School of Pathology and Laboratory Medicine, the University of Western Australia.
33. Institute of Social and Preventive Medicine (IUMSP), CHUV and Faculty of Biology and Medicine, Lausanne, Switzerland.
34. UCL Institute of Health Equity, Department of Epidemiology & Public Health, London, UK.
35. Instituto Medicina Molecular, Faculdade de Medicina Universidade de Lisboa, Lisbon, Portugal.
36. Servico Neurologia, Hospital de Santa Maria, Lisbon, Portugal.
37. Instituto Nacional de Saude Doutor Ricardo Jorge, Lisbon, Portugal.
38. Faculdade Ciencias Universidade Lisboa, Campo Grande, Lisbon, Portugal.
39. Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, University College London, London, UK.
40. Department of Cardiology, University Medical Center Groningen, Groningen, The Netherlands.
41. Research Centre for Prevention and Health, Capital Region of Denmark, Glostrup University Hospital, Glostrup, Denmark.

42. Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands
43. Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands.
44. Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands.
45. Vascular Screening and Diagnostic Centre, 2 Kyriakou Matsi str, Ayios Dometios, Nicosia, Cyprus.
46. Deparment of Vascular Surgery, Imperial College, London, UK.
47. Cyprus Cardiovascular Disease Educational and Research Trust, Nicosia, Cyprus.
48. South Asia Network for Chronic Disease, Public Health Foundation of India, New Delhi, India.
49. Department of Public Health & Caring Sciences, Uppsala University, Uppsala University Hospital, Uppsala, Sweden.
50. School of Surgery, University of Western Australia, Perth, Australia.
51. Stroke Unit, Department of Neurology, Royal Perth Hospital, 197 Wellington Street, Perth, Australia.
52. School of Medicine and Pharmacology, The University of Western Australia, Nedlands, Perth, Australia.
53. Centre for Cardiovascular Genetics, Institute of Cardiovascular Science, Rayne Building, University College London, London, UK.
54. Department of Vascular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands.
55. Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands.
56. Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK.
57. School of Population Health, University of South Australia, Adelaide, Australia.

58. School of Psychiatry & Clinical Neurosciences, University of Western Australia, Perth, Western Australia, Australia
59. Department of Psychiatry, Royal Perth Hospital, Perth, Western Australia, Australia.
60. Department of Primary Care and Public Health and Primary Care, University of Cambridge, Cambridge, UK.
61. Center for Molecular Medicine, Karolinska University Hospital Solna, Stockholm, Sweden.
62. The IMPROVE study group list of authors is as follows: D. Baldassarre (Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano, Milan, Italy and Centro Cardiologico Monzino, IRCCS, Milan Italy.); F. Veglia (Centro Cardiologico Monzino, IRCCS, Milan Italy); A. Hamsten (Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden.); S.E. Humphries (British Heart Foundation Laboratories, University College of London, Department of Medicine, Rayne Building, London, United Kingdom.); R. Rauramaa (Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland.); Ulf de Faire (Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, and Department of Cardiology, Karolinska University Hospital, Solna, Karolinska Institutet, Stockholm, Sweden.); A.J. Smit (Department of Medicine, University Medical Center Groningen, Groningen, the Netherlands.); P. Giral (Assistance Publique - Hopitaux de Paris; Service Endocrinologie-Metabolisme, Groupe Hôpitalier Pitie-Salpetriere, Unités de Prévention Cardiovasculaire, Paris, France.); S. Kurl (Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio Campus.); E. Mannarino (Internal Medicine, Angiology and Arteriosclerosis Diseases, Department of Clinical and Experimental Medicine, University of Perugia, Perugia, Italy.); E. Grossi (Bracco Milan Italy.); R. Paoletti (Dipartimento di

Scienze Farmacologiche e Biomolecolari, Università di Milano, Milan, Italy.); E. Tremoli (Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano, Milan, Italy and Centro Cardiologico Monzino, IRCCS, Milan Italy.)

63. Danish Cancer Society, Strandboulevarden, Copenhagen, Denmark.
64. National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark.
65. National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
66. Complex Genetics Section, Department of Medical Genetics (DBG), University Medical Center Utrecht, Utrecht, The Netherlands.
67. Dept. Pathology and Medical Biology, Medical Biology division, Molecular Genetics, University Medical Center Groningen and Groningen University, Groningen, The Netherlands.
68. Department of Primary Care & Population Health, UCL, Royal Free Campus, Rowland Hill St, London, UK.
69. Division of Population Health Sciences and Education, St George's, University of London, London, UK.
70. Instituto Gulbenkian Ciencia, Oeiras, Portugal.
71. Biofig - Center for Biodiversity, Functional and Integrative Genomics, Campus da FCUL, Lisboa, Portugal.
72. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
73. Department of Cardiovascular Medicine, University of Oxford, Oxford, UK.
74. Boston University, Boston, MA, USA.
75. National Heart, Lung, and Blood Institute's The Framingham Heart Study, Framingham, MA, USA.

76. The InterAct Consortium list of authors is as follows: C. Langenberg (MRC Epidemiology Unit, Cambridge, UK); S. Sharp (MRC Epidemiology Unit, Cambridge, UK); N.G. Forouhi (MRC Epidemiology Unit, Cambridge, UK); P.W. Franks (Lund University, Malmö, Sweden); M.B. Schulze (German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany); N. Kerrison (MRC Epidemiology Unit, Cambridge, UK); U. Ekelund (MRC Epidemiology Unit, Cambridge, UK); I. Barroso (Wellcome Trust Sanger Institute, Cambridge, UK); S. Panico (Federico II University, Naples, Italy); M.J. Tormo (Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain); J. Spranger (Charité University Berlin, Germany); S. Griffin (MRC Epidemiology Unit, Cambridge, UK); Y.T. van der Schouw (University Medical Center Utrecht, the Netherlands); P. Amiano (Public Health Division of Gipuzkoa, San Sebastian, Spain); E. Ardanaz (Navarre Public Health Institute, Pamplona, Spain); L. Arriola (Public Health Division of Gipuzkoa, San Sebastian, Spain); B. Balkau (INSERM, University Paris Sud, France); A. Barricarte (Epidemiology, Prevention and Promotion Health Service, Pamplona, Spain); J.W.J. Beulens (University Medical Center, Utrecht, the Netherlands); H. Boeing (German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany); H.B. Bueno-de-Mesquita (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); B. Buijsse (German Institute of Human Nutrition Potsdam-Rehbruecke, Germany); M.D. Chirlaque Lopez (Murcia Regional Health Authority, Spain); F. Clavel-Chapelon (INSERM, University Paris Sud, France); F.L. Crowe (University of Oxford, UK); B. de Lauzon-Guillan (INSERM, University Paris Sud, France); P. Deloukas (Wellcome Trust Sanger Institute, Cambridge, UK); M. Dorronsoro (Public Health Division of Gipuzkoa, San Sebastian, Spain); D. Drogan (German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany); P. Froguel (Imperial College London, UK); C. Gonzalez (Catalan Institute of Oncology, Barcelona, Spain); S. Grioni (Fondazione

IRCCS Istituto Nazionale Tumori, Milan, Italy); L. Groop (University Hospital Scania, Malmö, Sweden); C. Groves (University of Oxford, UK); P. Hainaut (International Agency for Research of Cancer, Lyon, France); J. Halkjaer (Danish Cancer Society, Copenhagen, Denmark); G. Hallmans (Umea University, Sweden); T. Hansen (Hagedorn Research Institute, Copenhagen, Denmark); J.M. Huerta Castaño (Murcia Regional Health Authority, Spain); R. Kaaks (German Cancer Research Centre, Heidelberg, Germany); T.J. Key (University of Oxford, UK); K.T. Khaw (University of Cambridge, UK); A. Koulman (MRC Human Nutrition Research, Cambridge, UK); A. Mattiello (Federico II University, Naples, Italy); C. Navarro (Murcia Regional Health Authority, Spain); P. Nilsson (Lund University, Malmö, Sweden); T. Norat (Imperial College London, UK); K. Overvad (School of Public Health, Aarhus, Denmark); L. Palla (MRC Epidemiology Unit, Cambridge, UK); D. Palli (Cancer Research and Prevention Institute (ISPO), Florence, Italy); O. Pedersen (Hagedorn Research Institute, Copenhagen, Denmark); P.H. Peeters (University Medical Center Utrecht, the Netherlands); J.R. Quirós (Asturias Health and Health Care Council, Oviedo, Spain); A. Ramachandran (India Diabetes Research Foundation, Chennai, India); L. Rodriguez-Suarez (Asturias Health and Health Care Council, Oviedo, Spain); O. Rolandsson (Umea University, Sweden); D. Romaguera (Imperial College London, UK); I. Romieu (International Agency for Research of Cancer, Lyon, France); C. Sacerdote (Center for Cancer Prevention, Torino, Italy); M.J. Sánchez (Andalusian School of Public Health, Granada, Spain); A. Sandbaek (School of Public Health, Aarhus, Denmark); N. Slimani (International Agency for Research of Cancer, Lyon, France); I. Sluijs (University Medical Center, Utrecht, the Netherlands); A.M.W. Spijkerman (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); B. Teucher (German Cancer Research Centre, Heidelberg, Germany); A. Tjonneland (Danish Cancer Society,

Copenhagen, Denmark); R. Tumino (Cancer Registry and Histopathology Unit, Ragusa, Italy); D.L. van der A (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); W.M.M. Verschuren (National Institute for Public Health and the Environment, Bilthoven, the Netherlands); J. Tuomilehto (University of Helsinki, Finland); E. Feskens (University of Wageningen, the Netherlands); M. McCarthy (University of Oxford, UK); E. Riboli (Imperial College London, UK); N.J. Wareham (MRC Epidemiology Unit, Cambridge, UK).

77. Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Utah, USA.
78. Schulich School of Medicine and Dentistry, University of Western Ontario, London, ON, Canada.
79. Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA,USA.
80. University of California San Francisco, CA, USA.
81. National Heart, Lung and Blood Institute Bethesda, MD, USA.
82. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.
83. Jackson Heart Study, Jackson State University, Jackson, MS, USA; School of Health Sciences, Department of Epidemiology and Biostatistics, Jackson State University, Jackson, MS, USA.
84. School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA.
85. Preventive Medicine and Epidemiology, Evans Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA.
86. Department of Laboratory Medicine and Pathology, University of Minnesota, USA.

87. Baylor College of Medicine, Department of Medicine, Division of Atherosclerosis & Vascular Medicine, Houston, Texas, USA.
88. Medical Genetics Institute, Department of Medicine, Cedars-Sinai, Los Angeles, USA.
89. Division of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Utah, USA.
90. Department of Medical Genetics, Biomedical Genetics, University Medical Center, Utrecht, The Netherlands.
91. British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK.
92. Genetics, R&D, GlaxoSmithKline, Stevenage, UK.
93. Beth Israel Deaconess Medical Center, Boston, MA, USA.
94. Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA.
95. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA.
96. Department of Genetics, University of North Carolina School of Medicine at Chapel Hill, Chapel Hill, North Carolina, USA.
97. College of Pharmacy, The University of New Mexico, Albuquerque, NM, USA.
98. The Copenhagen General Population Study, Herlev Hospital, Copenhagen, Denmark.
99. Faculty of Health Sciences, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark.
100. Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Denmark.
101. Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK.

102. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

Table 1. Comparison of linear and non-linear instrumental variable estimates for selected cardiovascular traits in the Alcohol-*ADH1B* Consortium.

Outcome	n	Linear IV approach		Non-linear IV approach						
				Mean LATE		LATE intercept		LATE slope		
		Estimate	95% CI[A]	Estimate	95% CI[A]	Estimate	95% CI[A]	Estimate	95% CI[A]	*P*[B]
SBP (mmHg)	78172	5.2	3.2, 7.3	4.9	2.6, 7.5	-2.2	-7.5, 3.4	3.3	1.0, 5.5	0.004
Non-HDL-C (mmol/L)	60140	0.13	-0.02, 0.28	0.25	0.06, 0.45	-0.54	-0.94, -0.120	0.37	0.19, 0.55	<0.001
HDL-C (mmol/L)	60227	-0.02	-0.07, 0.03	-0.01	-0.07, 0.06	-0.02	-0.15, 0.14	0.00	-0.06, 0.06	0.91
BMI (kg/m^2)	79454	0.7	0.2, 1.2	1.0	0.4, 1.5	-1.0	-2.5, 0.3	0.9	0.3, 1.4	0.002
WC (cm)	57172	2.8	1.3, 4.4	2.7	1.1, 4.5	-1.8	-5.8, 1.9	2.0	0.6, 3.6	0.01
CRP[C] (mg/l)	63367	0.17	0.03, 0.31	0.18	0.03, 0.38	-0.39	-0.77, 0.03	0.26	0.10, 0.43	0.001
IL-6[C] (pg/ml)	23535	0.30	0.16, 0.45	0.35	0.10, 0.53	0.10	-0.24, 0.85	0.13	-0.34, 0.29	0.41
TG[C] (mmol/L)	63667	0.01	-0.06, 0.07	0.01	-0.09, 0.07	0.04	-0.15, 0.21	-0.02	-0.10, 0.06	0.67

SBP, systolic blood pressure; Non-HDL-C, non-high density lipoprotein cholesterol; HDL-C, high density lipoprotein cholesterol; BMI, body mass index; WC, waist circumference; CRP, C-reactive protein; IL-6, interleukin 6; TG, triglycerides.

[A]Derived using 10,000 bootstrap samples. [B]Approximate Z-test using the bootstrap standard error. [C]Log transformed prior to analysis.

Table 2. Predicted difference in cardiovascular traits relative to zero alcohol consumption at several levels of alcohol consumption and predicted curve features in the Alcohol-*ADH1B* Consortium. Only calculated for traits with evidence of non-linearity.

Outcome	Difference in outcome (95% CI[A])				Level of alcohol consumption at nadir (units/week[C]) (95% CI[A])	Difference in outcome at optimal alcohol consumption(95% CI[A])	Level of alcohol consumption with outcome equal to that at zero (units/week[C]) (95% CI[A])
	3.04 units/week[C]	12.15 units/week[C]	31.90 units/week[C]	84.52 units/week[C]			
SBP (mmHg)	0.1 (-5.5, 6.1)	5.2 (-2.6, 13.9)	12.4 (3.4, 22.1)	22.8 (12.2, 34.6)	1.0 (0.0, 3.6)	-0.7 (-5.4, 0.0)	2.8 (0.0, 19.6)
Non-HDL-C (mmol/L)	-0.39 (-0.79, 0.06)	-0.15 (-0.72, 0.47)	0.40 (-0.28, 1.10)	1.30 (0.45, 2.16)	3.2 (0.7, 6.0)	-0.39 (-0.85, -0.03)	16.9 (2.1, 48.2)
BMI (kg/m^2)	-0.6 (-2.2, 0.8)	0.2 (-2.0, 2.1)	1.6 (-0.8, 3.8)	3.9 (1.2, 6.3)	2.3 (0.0, 6.0)	-0.6 (-2.3, 0.0)	10.1 (0.0, 48.4)
WC (cm)	-0.6 (-4.7, 3.5)	1.9 (-3.9, 7.8)	5.7 (-0.6, 12.5)	11.5 (4.5, 19.2)	1.5 (0.0, 5.4)	-0.8 (-4.9, 0.0)	5.3 (0.0, 37.4)
CRP[B] (mg/l)	-0.29 (-0.68, 0.15)	-0.15 (-0.68, 0.50)	0.22 (-0.37, 0.95)	0.83 (0.15, 1.69)	3.5 (0.0, 7.2)	-0.30 (-0.75, 0.00)	19.4 (0.0, 66.0)

SBP, systolic blood pressure; Non-HDL-C, non-high density lipoprotein cholesterol; BMI, body mass index; WC, waist circumference; CRP, C-reactive protein.

[A]Derived using 10,000 bootstrap samples. [B]Log transformed prior to analysis. [C]Weekly units of alcohol values are medians of observed values in categories representing low (1-7 units/week), moderate (7-21 units/week), heavy (21-70 units/week) and very heavy (70+ units/week) alcohol consumption in the analysis of Holmes et al.[13]

Box 1. Summary of proposed method for testing for a non-linear causal effect.

1. For the observed data and for each of *K* bootstrap samples:
   1.1. Regress *X* on *G* for all subjects, giving estimated regression coefficient $\hat{\beta}_{XG}$
   1.2. Discretise *X* into units of $\hat{\beta}_{XG}$, that is derive the discrete variable $X^* = \left\lfloor \frac{X}{\hat{\beta}_{XG}} \right\rfloor$
   1.3. For each discrete value of *j*:
      1.3.1. Regress *Y* on X^* using only the subjects for which $X_i^* = j$ and $G_i = 1$, or $X_i^{*'} = j - 1$ and $G_i = 0$. Among these subjects there is no variation in X^* that is not explained by *G*.
      1.3.2. This yields $\hat{\tau}_j$, the estimated local average treatment effect (LATE) for level *j* of X^*
      1.3.3. Rescale $\hat{\tau}_j$ by $\hat{\beta}_{XG}$ to the original scale of *X*
   1.4. Obtain the mean LATE by fixed-effects meta-analysis of $\hat{\tau}_j$
   1.5. Meta-regress $\hat{\tau}_j$ on *j* to obtain the intercept and slope of the LATEs, corresponding to a quadratic causal model
2. Obtain empirical confidence intervals on the mean LATE and the LATE intercept and slope from the bootstrap samples

Figure legends

Figure 1. Directed acyclic graphs encoding a) the standard Mendelian randomisation assumptions: (i) G is associated with X, ii) G is not associated with confounders U of the X-Y association, and iii) G affects Y only via its association with X; b) how these are assumptions are affected by the discretisation of X in the proposed non-linear Mendelian randomisation approach.

Figure 2. Local average treatment effects (LATEs) of log(weekly units of alcohol + 1) on systolic blood pressure. Circular markers are LATEs; bars are 95% pointwise confidence intervals; dashed line is estimated mean LATE; solid line is estimated linear LATE trend; dotted line is linear IV estimate using the ratio method (virtually indistinguishable from the estimated mean LATE).

Figure 3. Predicted difference in non-high density lipoprotein cholesterol (non-HDL-C) relative to zero alcohol consumption across the range of values of observed alcohol consumption, with estimated optimal level of alcohol consumption (3.2 (95% confidence interval (CI): 0.7, 6.0) units/week), estimated difference in non-HDL-C relative to zero alcohol consumption at optimal level (-0.39 (95% CI: -0.85, -0.03) mmol/L), and estimated level of alcohol consumption with the same level of non-HDL-C as at zero (16.9 (95% CI: 2.1, 48.2) units/week) indicated.

References

1. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008; **5**: e177.
2. Jansen H, Samani NJ, Schunkert H. Mendelian randomization studies in coronary artery disease. *Eur Heart J* 2014; **35**: 1917-24.
3. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. In press.
4. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; **27**: 1133-63.
5. Angrist JD, Pischke J-S. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *J Econ Perspectives* 2010; **24**: 3-30.
6. Vukcevic D, Hechter E, Spencer C, Donnelly P. Disease model distortion in association studies. *Genet Epidemiol* 2011; **35**: 278-90.
7. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ* 2011; **342**: d671.
8. Mukamal KJ, Mackey RH, Kuller LH, et al. Alcohol consumption and lipoprotein subclasses in older adults. *J Clin Endocrinol Metab* 2007; **92**: 2559-66.
9. Greenfield JR, Samaras K, Jenkins AB, Kelly PJ, Spector TD, Campbell LV. Moderate alcohol consumption, estrogen replacement therapy, and physical activity are associated with increased insulin sensitivity: is abdominal adiposity the mediator? *Diabetes Care* 2003; **26**: 2734-40.

10. Imhof A, Froehlich M, Brenner H, Boeing H, Pepys MB, Koenig W. Effect of alcohol consumption on systemic markers of inflammation. *Lancet* 2001; **357***:* 763–7.
11. Albert MA, Glynn RJ, Ridker PM. Alcohol consumption and plasma concentration of C-reactive protein. *Circulation* 2003; **107**: 443-7.
12. Mukamal KJ, Jadhav PP, D'Agostino RB, et al. Alcohol consumption and hemostatic factors: analysis of the Framingham Offspring Cohort. *Circulation* 2001; **104**: 1367–73.
13. Holmes MV, Dale CE, Zuccolo L, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ* 2014; **349**: g4164.
14. Jackson R, Broad J, Connor J, Wells S. Alcohol and ischaemic heart disease: probably no free lunch. *Lancet* 2005; **366**: 1911-2.
15. Emberson JR, Shaper AG, Wannamethee SG, Morris RW, Whincup PH. Alcohol intake in middle age and risk of cardiovascular disease and mortality: accounting for intake variation over time. *Am J Epidemiol* 2005; **161**: 856-63.
16. Bergmann MM, Rehm J, Klipstein-Grobusch K, et al. The association of pattern of lifetime alcohol use and cause of death in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Int J Epidemiol* 2013; **42**: 1772-90.
17. Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. *Nat Rev Genet* 2005; **6**: 521-32.
18. Tolstrup JS, Nordestgaard BG, Rasmussen S, Tybjaerg-Hansen A, Gronbaek M. Alcoholism and alcohol drinking habits predicted from alcohol dehydrogenase genes. *Pharmacogenomics J* 2008; **8**: 220-7.
19. Zuccolo L, Fitz-Simon N, Gray R, et al. A non-synonymous variant in ADH1B is strongly associated with prenatal alcohol use in a European sample of pregnant women. *Hum Mol Gene* 2009; **18**: 4457-66.

20. Lawlor DA, Nordestgaard BG, Benn M, Zuccolo L, Tybjaerg-Hansen A, Davey Smith G. Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study. *Eur Heart J* 2013; **34**: 2519-28.
21. Newey WK, Powell JL. Instrumental variable estimation of nonparametric models. *Econometrica* 2003; **71**: 1565–78.
22. Chernozhukov V, Hansen C. An IV model of quantile treatment effects. *Econometrica* 2005; **73**: 245–61.
23. Horowitz JL. Applied nonparametric instrumental variables estimation. *Econometrica* 2011; **79**: 347–94.
24. Li D, Zhao H, Gelernter J. Strong association of the alcohol dehydrogenase 1B gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. *Biol Psychiatry* 2011; **70**: 504-12.
25. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62**: 467-75.
26. Angrist DA, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *J Amer Statist Assoc* 1995; **90**: 431-22.
27. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. 2011.
28. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol* 2011; **40**: 740-52.
29. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* 2012; **21**: 223-42.

Testing for non-linear causal effects using a binary genotype in a Mendelian randomisation study: application to alcohol and cardiovascular traits

Richard J Silverwood, Michael V Holmes, Caroline E Dale, Debbie A Lawlor, John C Whittaker, George Davey Smith, David A Leon, Tom Palmer, Brendan J Keating, Luisa Zuccolo, Juan P Casas, Frank Dudbridge, on behalf of the Alcohol-*ADH1B* Consortium

Supplementary material

Simulations – Methods

We conducted simulations to assess the proposed approach in terms of bias and coverage at a variety of sample sizes and under different *X*-*Y* associations. We performed additional simulations to explore the effects of β_{XG} heterogeneity and *G*-*U* interaction at both the individual and subgroup level.

The simulation parameters were chosen to resemble those observed in the Alcohol-*ADH1B* Consortium, and a plausible degree of confounding was applied. In each simulation the data were generated according to

$$Y \sim N(\beta_1 X + \beta_2 X^2 - 2U + U^2, 5^2)$$

with the following specifications remaining constant:

$G \sim \text{Bin}(1, 0.3); U \sim N(0,1)$

Thus, we allow for a quadratic causal effect of *X* on *Y*, but also for quadratic effects of confounders. The following simulations were conducted:

a) *Sample size*. The effect of sample size was examined using the null model with $(\beta_1, \beta_2) = (0,0)$. Sample size was varied within $\{5000, 10000, 50000, 100000\}$. *X* was simulated using $X \sim N(-0.2G + 0.5U, 1)$.

b) *X-Y associations.* A variety of *X-Y* associations were explored using a sample size of 10000. Linear associations were generated using $(\beta_1, \beta_2) = \{(-5,0), (-2,0), (-1,0), (1,0), (2,0), (5,0)\}$; quadratic associations were generated using $(\beta_1, \beta_2) = \{(2,-5), (1,-2), (2,-1), (-2,1), (-1,2), (-2,5)\}$. *X* was again simulated using $X \sim N(-0.2G + 0.5U, 1)$.

c) *Individual-level β_{XG} heterogeneity*. The key assumption of our approach is that β_{XG} is constant for all individuals. The degree of variability introduced into these simulations regarding individual-level β_{XG} heterogeneity was informed by the observation that within the *ADH1B* Consortium the study-specific estimates of the association between *rs1229984* and log-alcohol had mean -0.235 and SD 0.121, along with the assumption that individual-level heterogeneity is likely to be somewhat greater than this. We defined *X* using $X \sim N((-0.2 + \alpha)G + 0.5U, 1)$ where $\alpha \sim N(0, \gamma^2)$ was simulated independently for each subject within a sample size of 10,000 and γ varied within $\{0, 0.1, 0.2, 0.3, 0.5, 1\}$. Thus, for $\gamma = 0.1$ β_{XG} is expected to lie between -0.4 and 0 in 95% of subjects, and approximately 61% of subjects are

expected to be compliers in the sense that they are moved by the instrument to the adjacent bin of size $\beta_{XG} = -0.2$. Simulations were conducted using the quadratic data generating model with $(\beta_1, \beta_2) = (-1, 2)$.

d) *Individual-level G-U interaction*. We defined *X* using $X \sim N(-0.2G + 0.5U + \delta GU, 1)$ where δ varied within $\{0, 0.1, 0.2, 0.3, 0.5, 1\}$ in a sample of size 10,000. This means that the effect of *G* on *X* is allowed to vary by levels of the confounder *U*. Simulations were conducted using the quadratic data generating model with $(\beta_1, \beta_2) = (-1, 2)$.

e) *β_{XG} heterogeneity between unknown subgroups.* The degree of variability introduced into these and subsequent simulations regarding β_{XG} heterogeneity between known or unknown subgroups were informed by the observation that within the *ADH1B* Consortium the study-specific estimates of the association between *rs1229984* and log-alcohol had mean -0.235 and SD 0.121. We defined *X* using $X \sim N((-0.2 + \alpha)G + 0.5U, 1)$ where $\alpha \sim N(0, \gamma^2)$ took the same value within subgroups of 1000 subjects within an overall sample size of 10,000 and γ varied within $\{0, 0.01, 0.02, 0.03, 0.05, 0.1\}$. Thus, for $\gamma = 0.1$ β_{XG} is expected to lie between --0.4 and 0.0 in 95% of subgroups. Simulations were conducted using the quadratic data generating model with $(\beta_1, \beta_2) = (-1, 2)$. Data were analysed as previously described, with all subgroups pooled together.

f) *β_{XG} heterogeneity between known subgroups. X* was again defined using $X \sim N((-0.2 + \alpha)G + 0.5U, 1)$ where $\alpha \sim N(0, \gamma^2)$ now took the same value within subgroups of 10,000 subjects within an overall sample size of 100,000 and γ varied

between $\{0, 0.01, 0.02, 0.03, 0.05, 0.1\}$. Simulations were again conducted using the quadratic data generating model with $(\beta_1, \beta_2) = (-1, 2)$. Data were analysed with all subgroups pooled together and adjustment for subgroup using indicator variables..

g) *G-U interaction which varies between unknown subgroups.* In addition to individual-level *G-U* interaction, the degree of interaction may also vary between subgroups in a population. We investigated this issue in the situation where the subgroups between which the *G-U* interaction varied are unknown. We defined *X* using $X \sim N(-0.2G + 0.5U + \delta GU, 1)$ where $\delta \sim N(0, \gamma^2)$ took the same value within subgroups of 1000 subjects within an overall sample size of 10,000 and γ varied within $\{0, 0.01, 0.02, 0.03, 0.05, 0.1\}$. Simulations were conducted using the quadratic data generating model with $(\beta_1, \beta_2) = (-1, 2)$. Analysis proceeded as previously described, with data from all subgroups pooled together.

When regressing *Y* on X^* we only estimated a LATE if there were at least 5 subjects with $X_i^* = j$ and $G_i = 1$ and 5 subjects with $X_i^* = j - 1$ and $G_i = 0$, as regressions using smaller numbers of subjects were found to lead to anomalous estimated LATEs with spuriously high precision. A thousand simulations of each specification were conducted, with percentile bootstrap confidence intervals derived using 1000 bootstrap samples within each simulation. Standard 2SLS and quadratic OLS estimates were also calculated for comparison.

The analysis was conducted using R version 2.13.1

Simulations - Results

a) *Sample size*. Results for varying sample size are presented in Figure S1. The linear IV estimate, mean LATE, LATE intercept and LATE slope were unbiased even with a sample size of 5000. The variability in the mean LATE and LATE slope was similar to that of the linear IV estimate. The quadratic OLS estimates were significantly biased at all sample sizes. The corresponding coverages are shown in Figure S2. The linear IV estimate, mean LATE and LATE intercept coverages varied between 94% and 96%, even at smaller sample sizes. The coverage of the LATE slope was somewhat conservative (around 97%) at smaller sample sizes.

b) *X-Y associations.* In the linear data generating models the linear IV estimate, mean LATE, LATE intercept and LATE slope were all unbiased (Figure S3). The mean LATE and LATE slope again had similar variability to the linear IV estimate. The quadratic OLS estimates were biased in all scenarios. The coverages of the linear IV estimate, mean LATE, and LATE intercept were generally between 94% and 96%, but the coverage of the LATE slope was approximately 96% across the range of linear coefficients (Figure S4). This suggests that our procedure gives a slightly conservative test of no non-linear effect. In the quadratic data generating models the LATE intercepts and LATE slopes were unbiased for all combinations of coefficients but the quadratic OLS estimates were again heavily biased (Figure S5). The coverages of the LATE intercept and LATE slope were generally between 94% and 96%, though the LATE slope coverage was as high as 97.5% for $(\beta_1, \beta_2) = (-2, 5)$ (Figure S6).

c) *Individual-level* β_{XG} *heterogeneity*. For small values of γ the LATE intercept and LATE slope displayed little bias (Figure S7). However, for values of γ of 0.2 and greater noticeable bias began to appear, particularly in the LATE slope. By $\gamma = 1$ the

bias was so great that the 95% range of estimates of the LATE slope did not include the target value. Both the LATE intercept and LATE slope had acceptable levels of coverage for values of γ up to 0.3 (Figure S8). For values of γ larger than 0.5 the coverage of the LATE slope was drastically reduced. However, it should be noted that the larger values of γ in these simulations do represent very extreme cases (e.g. for $\gamma = 1$ β_{XG} is expected to lie between -2.2 and 1.8 in 95% of subjects, so in almost 50% of subjects the gene would have an effect on alcohol consumption in the opposite direction to that assumed). Thus, given the reliance of our method on the assumption of β_{XG} homogeneity, the observed bias and reduced coverage is not unexpected.

d) *Individual-level G-U interaction*. Bias was observed for all values of δ greater than zero for both the LATE intercept and the LATE slope (Figure S9). The bias in the LATE intercept increased as δ increased, but the bias in the LATE slope decreased slightly at $\delta = 1$. The coverage of the parameters correspondingly generally deviated further from 95% as δ increased (Figure S10), reaching 74% for the LATE intercept and 59% for the LATE slope at $\delta = 0.5$. Coverage for the LATE intercept reduced further to 14% at $\delta = 1$, but improved slightly for the LATE slope.

e) *β_{XG} heterogeneity between unknown subgroups.* At no values of γ was there any appreciable bias in either the LATE intercept or the LATE slope (Figure S11). For all values of γ the LATE intercept had coverage close to 95%, but there was again some evidence that the coverage of the LATE slope was slightly conservative (Figure S12).

f) *β_{XG} heterogeneity between known subgroups.* At no values of γ was there any evidence of bias in either the LATE intercept or the LATE slope (Figure S13).

Coverage was approximately 96% for the LATE intercept and 97-99% for the LATE slope (Figure S14).

g) *G-U interaction which varies between unknown subgroups.* There was no bias in either parameter at $\gamma = 0.1$, but bias in both parameters, most noticeably in the LATE slope, increased as γ increased further (Figure S15). Coverage was approximately appropriate for the LATE intercept as far as $\gamma = 0.5$, though it was <90% for the LATE slope from $\gamma = 0.3$ onwards (Figure S16). By $\gamma = 1$ the coverages of the LATE intercept and slope were 84% and 23% respectively.

These results suggest that the LATE estimates are essentially unbiased with generally good coverage properties under null, linear and quadratic models. Reasonable levels of individual-level heterogeneity in β_{XG} were not found to lead to significant bias in the estimates. Low levels of between-subgroup heterogeneity in β_{XG} were also not found to lead to significant bias, whether or not the heterogeneity was adjusted for in the analysis. High levels of interaction between G and U led to bias in the estimates, but such interactions may be unlikely in practice. For example, in simulation d) a δ of 0.2 biased the LATE intercept upwards by 30% and the LATE slope upwards by 16%. This level of interaction means that the effect of G on X can be expressed as $-0.2 + 0.2U$. As $U \sim N(0,1)$, 95% of the values of U will lie between -2 and 2 (approximately). Thus the effect of G on X will lie between -0.6 and 0.2 approximately 95% of the time, depending on the value of U. Such a wide range of genetic effects, including a reversal of sign, may well be deemed implausible.

Overall, our simulations indicate that the LATE method is a useful extension to standard approaches in the non-linear setting.

Multiple studies

When data from multiple studies are available we must decide whether to estimate LATEs within each study, combining them to draw an overall conclusion, or to estimate LATEs on the combined data. In general, weak instrument bias is reduced by performing the MR analysis on the combined data.2 But because our approach relies on the genetic effect size β_{XG} being constant for all individuals, we need to consider whether this assumption is tenable across studies as well as within each study.

If β_{XG} varies across studies then clearly we should perform the procedure separately in each study. This however raises two problems. Firstly, the restriction to subjects having particular genotype-exposure combinations in each bin of X^* leads to small sample sizes for estimating some LATEs, leading to large standard errors on some $\hat{\tau}_j$ and considerable uncertainty on the final inference of non-linearity. This problem occurs particularly when one genotype is rare, or when β_{XG} is small, leading to narrow bin definitions, both of which apply to the Alcohol-*ADH1B* Consortium. Secondly, different β_{XG} across studies leads to different bin sizes across studies and different local causal effects being estimated. It is not clear how such effects should be combined into an overall inference on non-linearity.

For these reasons we assumed that β_{XG} is constant within and between studies and performed simulations to assess robustness to that assumption. As in the standard IV analysis, covariates for study, age and sex were included in both the estimation of β_{XG} and the LATEs in order to reduce the potential for confounding of β_{XG} and the LATEs.

Table S1. Design and genotyping characteristics of the studies included in the analysis.

Study	Study design	Sampling Frame	Number with DNA in this analysis	Number contributing to one or more of our analyses[A]	Year of blood sampling used for DNA extraction	Genotyping method	Country	HWE P value (exact significance probability)	Call rate (%)
ARIC	Cohort	Community	9557	9532	1987-89	IBC 50k CardioChip	USA	0.705	97.8
BRHS	Cohort	General practices	3843	3789	1998-2000	KASPar	UK	0.42	100
BWHHS	Cohort	General practices	3412	3407	1999-2001	Illumina HumanCVD array	UK	0.912	99.7
CaPS	Cohort	Electoral register & General practices	1102	1061	1993-1994	KASPar	UK	0.460	98.4
CARDIA	Cohort	Community	1433	1433	1995-1996	IBC 50k CardioChip	USA	4.97E-04	97.3
CCHS	Cohort	Population	9081	8985	1991-94	Nanogen	Denmark	0.522	99.6
CHS	Cohort	Community	3936	3919	1992–1993	IBC 50k CardioChip	USA	0.001	97.9
CYPRUS	Cohort	Community	730	729	2003-2008	TaqMan	Cyprus	0.081	99.9
Czech post-MONICA	Cohort	Administrative districts	2558	2555	2000-2001	PCR-RFLP	Czech Republic	0.801	97.9
DCH	Nested case cohort	General population (born in Denmark)	2736	2735	1993-97	TaqMan	Denmark	0.203	91.8
EAS	Cohort	General practices	873	873	2004	TaqMan	UK	0.693	95.6

ELSA	Cohort	Respondents of HSE	5450	5449	2004	KASPar	UK	0.263	98.8
EPIC Turin	Cohort	Population (Torino area)	4526	4314	2008	TaqMan	Italy	0.362	99
FHS	Cohort	Community	1082	312	1948-present	IBC 50k CardioChip	USA	0.002	99
HAPIEE Czech	Cohort	City districts	6678	6553	2003-2005	KASPar	Czech Republic	0.745	98.6
Inter99	RCT	Population	6332	6025	1999-2001	KASPar	Denmark	6.16E-27	97.6
Izhevsk	Case control	Population-based controls from CC	653	642	2008-2009	PCR + electrophoresis	Russia	0.192	>99
MESA	Cohort	Population	2293	2054	2000-2002	IBC 50k CardioChip	USA	0.012	97
NPHS II	Cohort	General practices	2659	2659	2000	TaqMan	UK	0.874	96.1
ULSAM	Cohort	General population (Uppsala County)	453	421	2004	Illumina Golden Gate	Sweden	0.775	98.91
Whitehall II	Cohort	Workplace (civil servants)	5029	4990	2002-2004	IBC 50k CardioChip	UK	0.106	99.3
WHI	Nested case control	Community	7882	7620	1993-1998	IBC 50k CardioChip	USA	3.15E-25	99.2

[A]Number of study members non-missing for weekly volume of alcohol, *rs1229984* polymorphism in *ADH1B*, age, sex, and one or more of the outcomes (systolic blood pressure, non-high density lipoprotein cholesterol, high density lipoprotein cholesterol, body mass index, waist circumference, C-reactive protein, interleukin 6 and triglycerides).

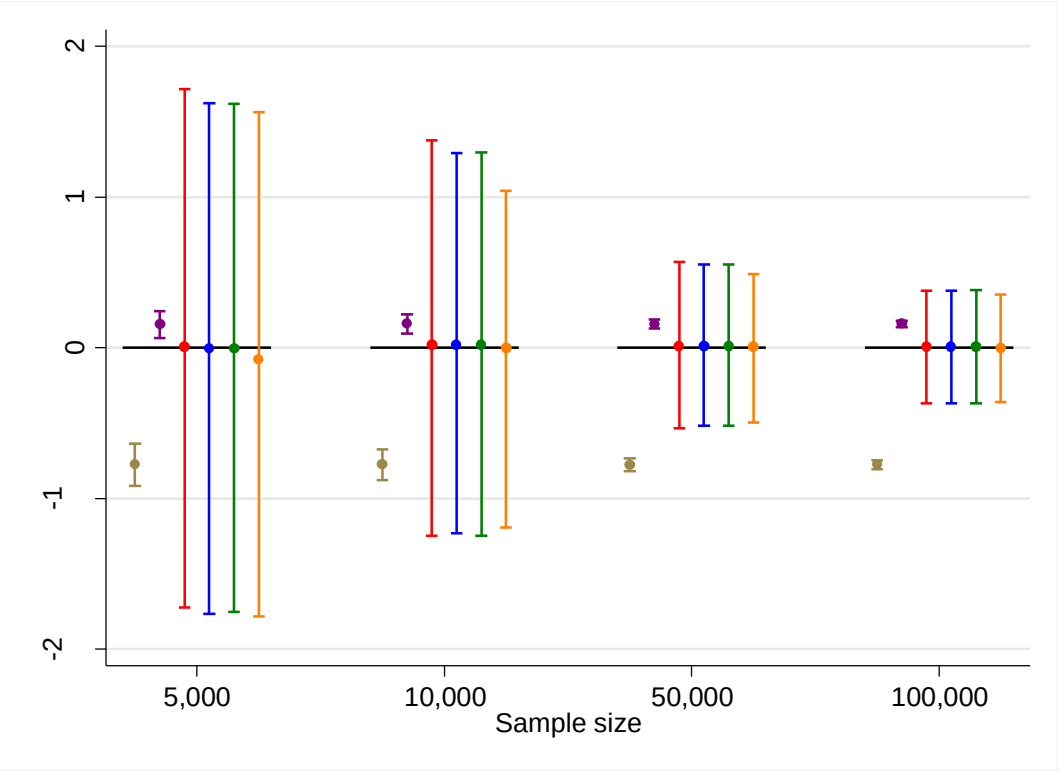

Figure S1. Observational quadratic model linear parameter estimates (brown) and quadratic parameter estimate (purple), two-stage least squares estimates (red), mean local average treatment effects (LATEs) (blue), LATE intercepts (green) and LATE slopes (orange) for different sample sizes in the null data generating model. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

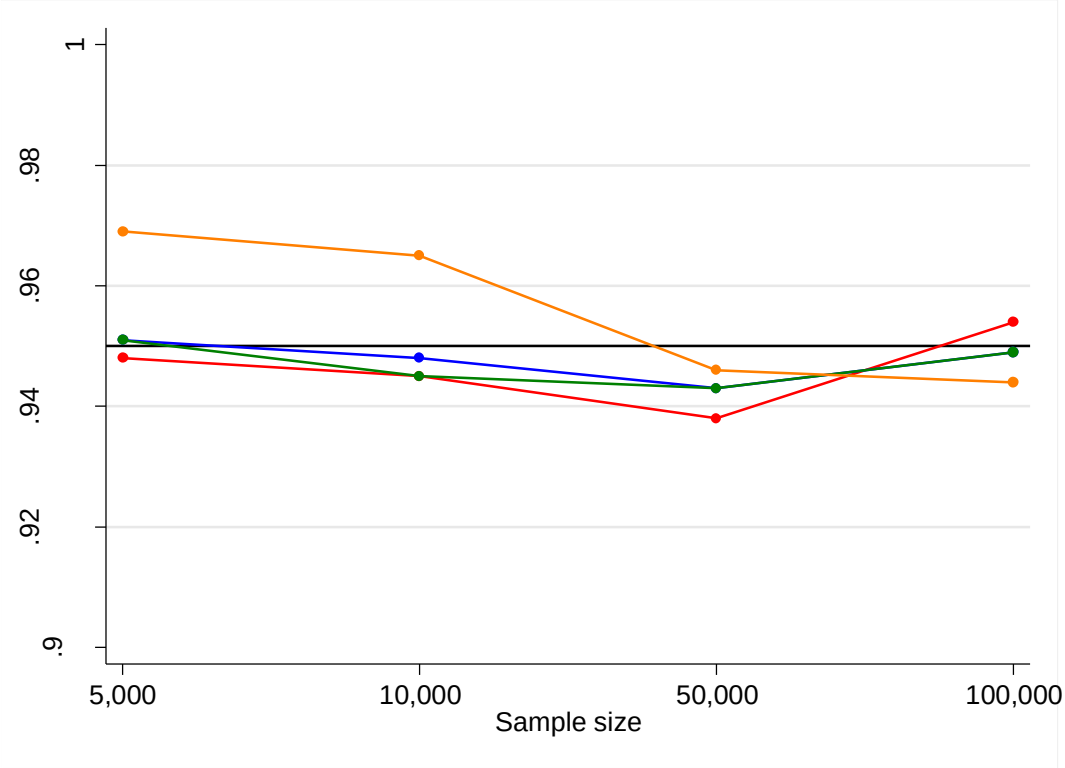

Figure S2. 95% confidence interval coverage for the two-stage least squares estimates (red), mean LATEs (blue), LATE intercepts (green) and LATE slopes (orange) for different sample sizes in the null data generating model. Horizontal solid line represents the target value (95%).

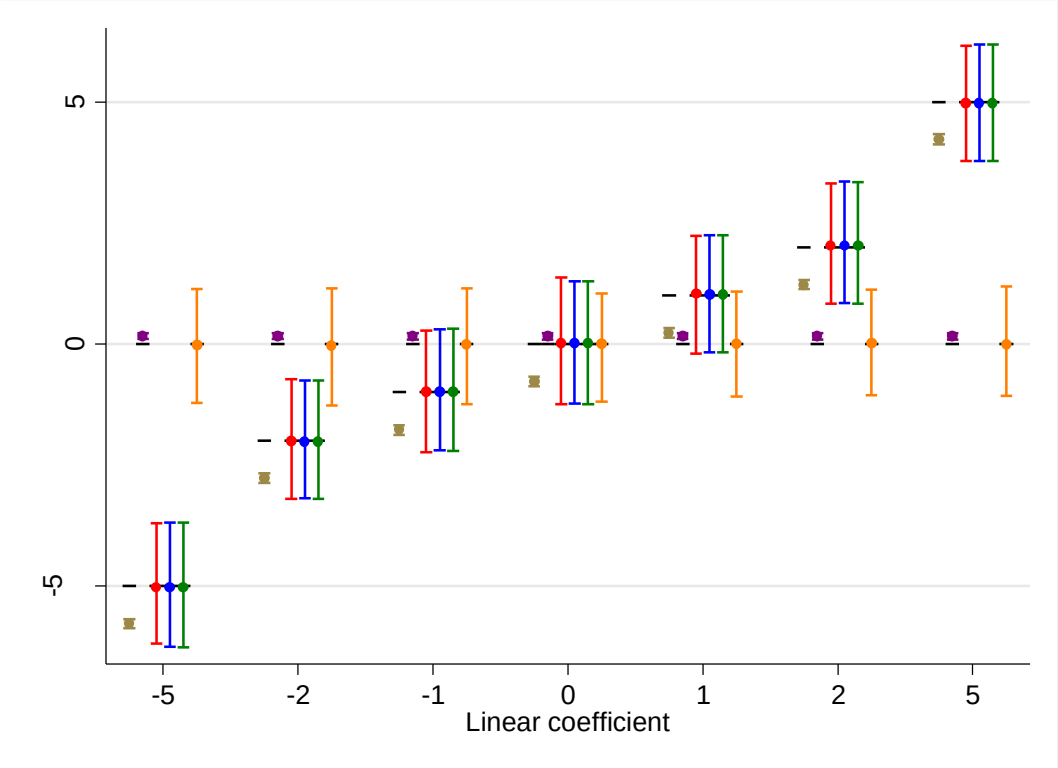

Figure S3. Observational quadratic model linear parameter estimates (brown) and quadratic parameter estimate (purple), two-stage least squares estimates (red), mean LATEs (blue), LATE intercepts (green) and LATE slopes (orange) for different linear coefficients in the linear data generating model. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

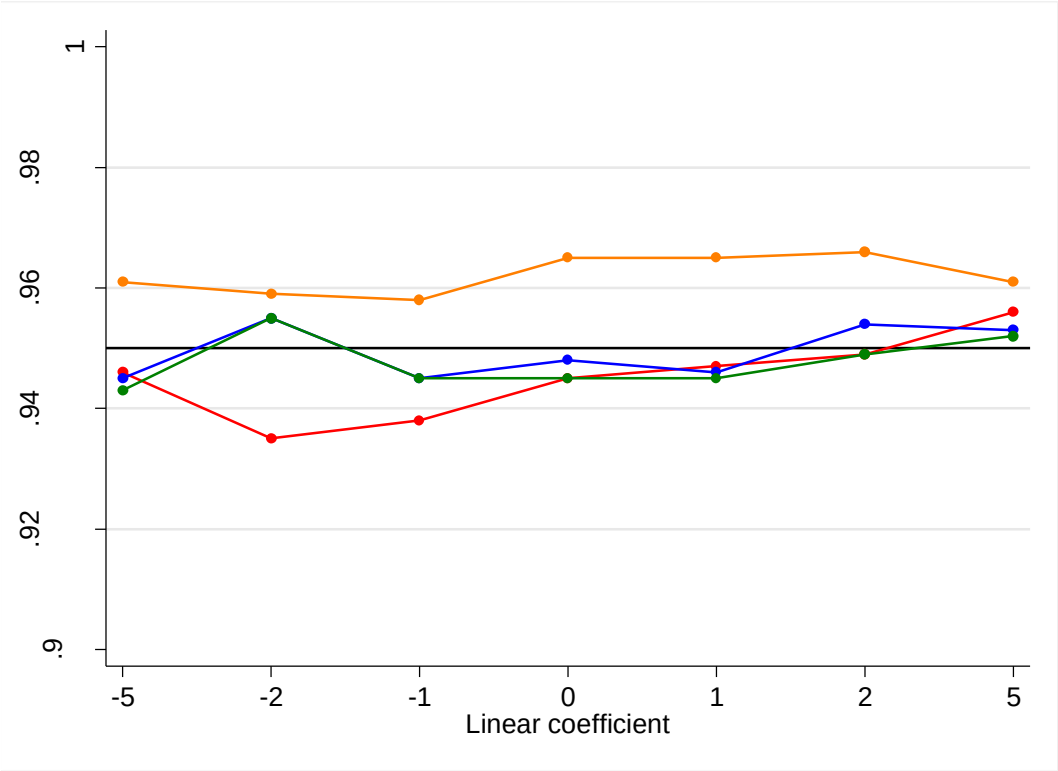

Figure S4. 95% confidence interval coverage for the two-stage least squares estimates (red), mean LATEs (blue), LATE intercepts (green) and LATE slopes (orange) for different linear coefficients in the linear data generating model. Horizontal solid line represents the target value (95%).

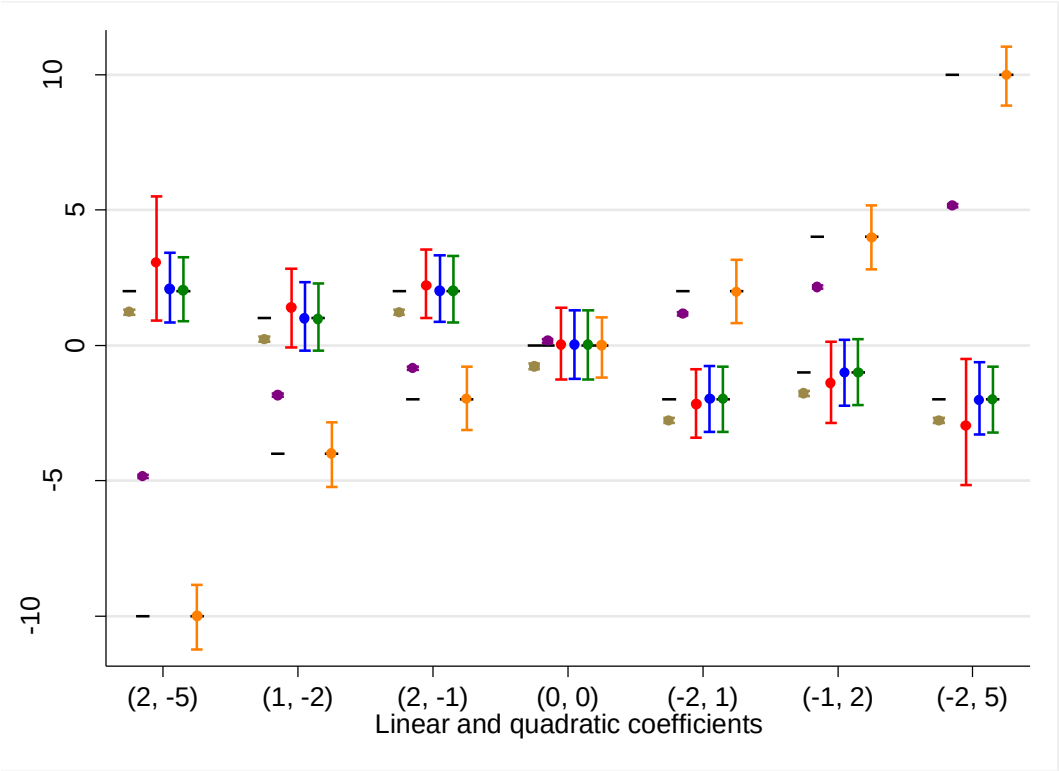

Figure S5. Observational quadratic model linear parameter estimates (brown) and quadratic parameter estimate (purple), two-stage least squares estimates (red), mean LATEs (blue), LATE intercepts (green) and LATE slopes (orange) for different combinations of linear and quadratic coefficients in the quadratic data generating model. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

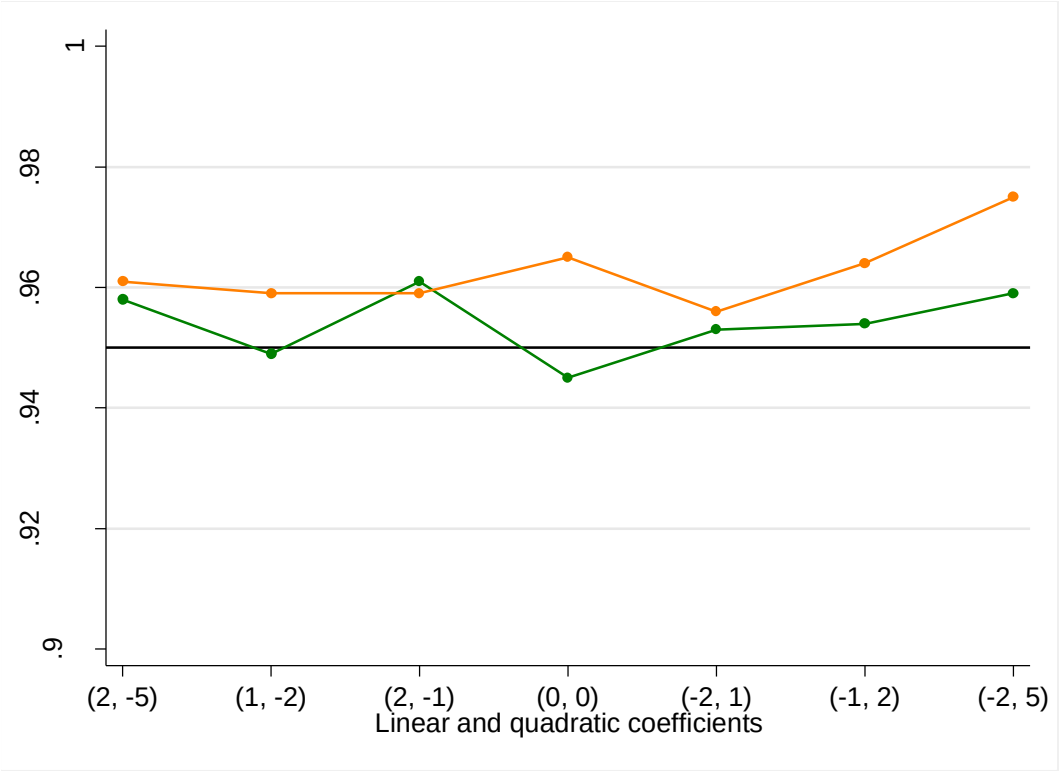

Figure S6. 95% confidence interval coverage for the LATE intercepts (green) and LATE slopes (orange) for different combinations of linear and quadratic coefficients in the quadratic data generating model. Horizontal solid line represents the target value (95%).

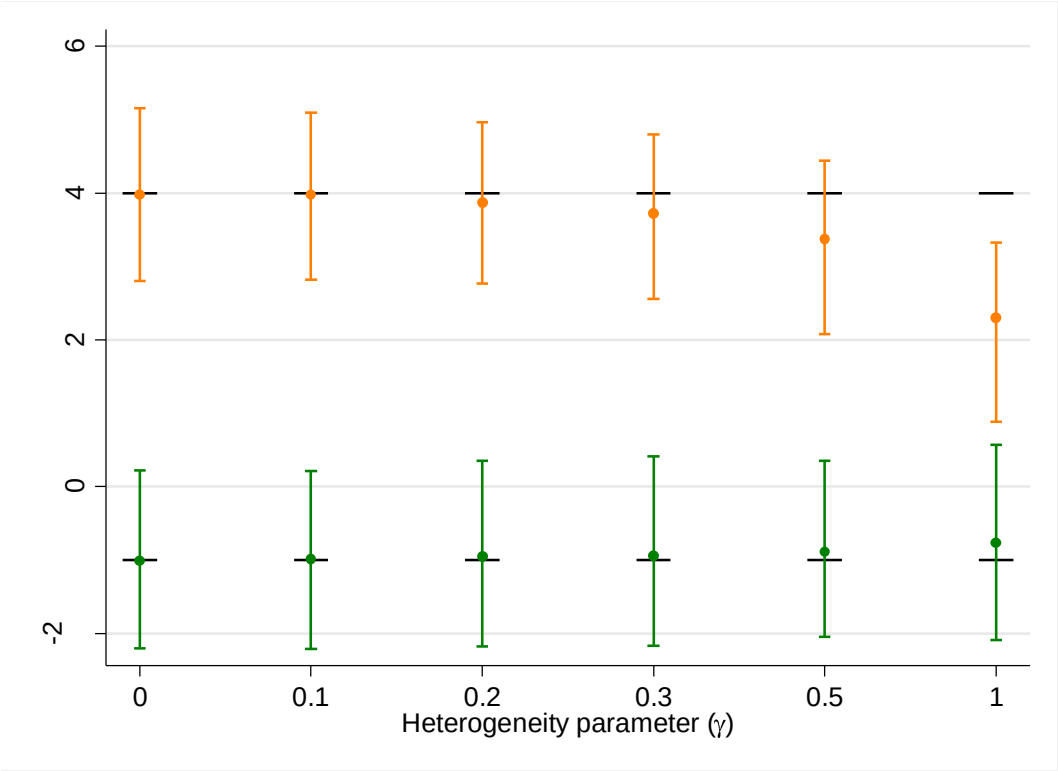

Figure S7. LATE intercepts (green) and LATE slopes (orange) for different degrees of individual-level heterogeneity of β_{xg} in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

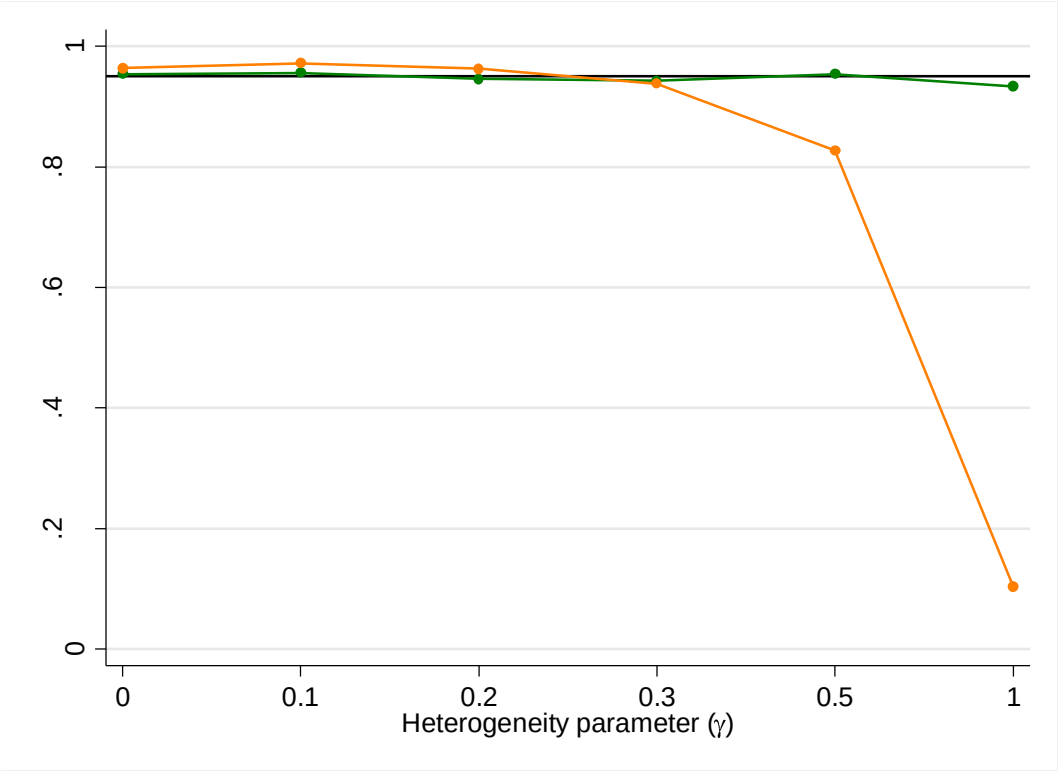

Figure S8. 95% confidence interval coverage for LATE intercepts (green) and LATE slopes (orange) for different degrees of individual-level heterogeneity of β_{xg} in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Horizontal solid line represents the target value (95%).

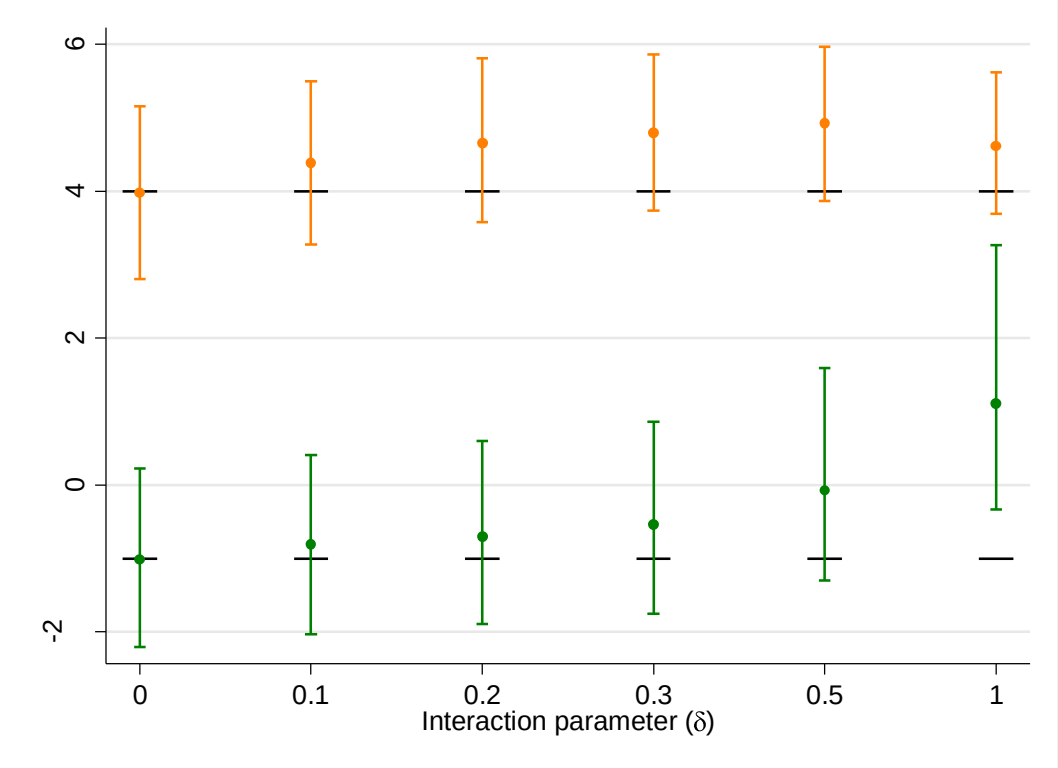

Figure S9. LATE intercepts (green) and LATE slopes (orange) for different degrees of individual-level G-U interaction in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

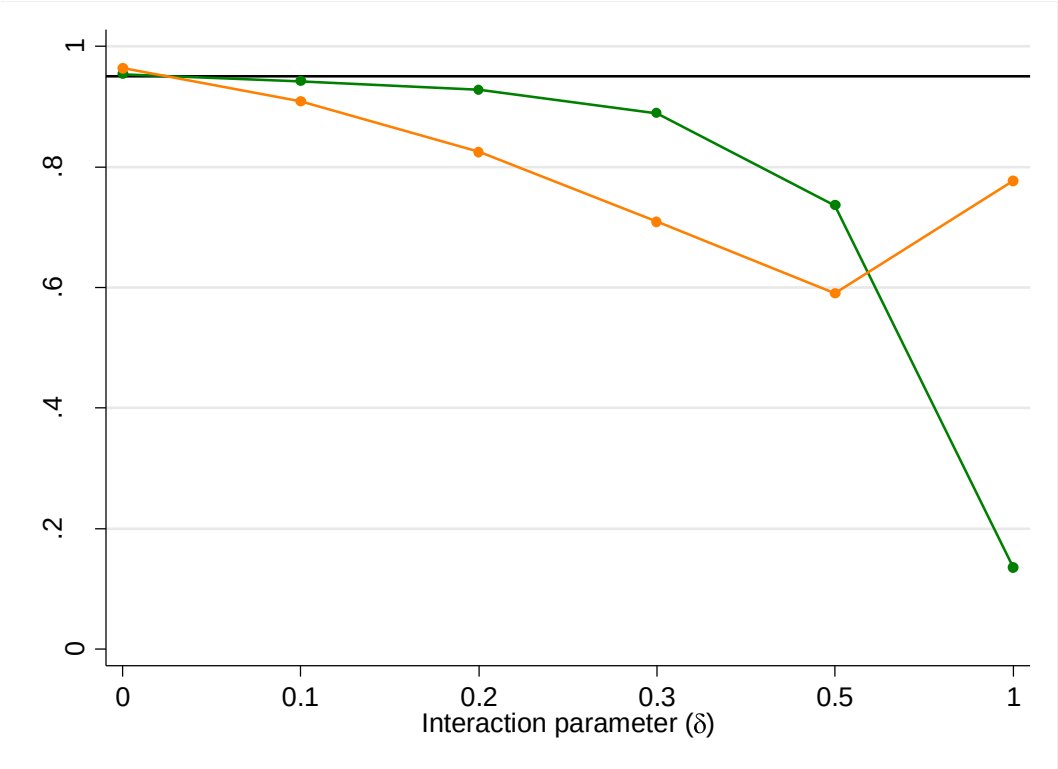

Figure S10. 95% confidence interval coverage for LATE intercepts (green) and LATE slopes (orange) for different degrees of individual-level G-U interaction in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Horizontal solid line represents the target value (95%).

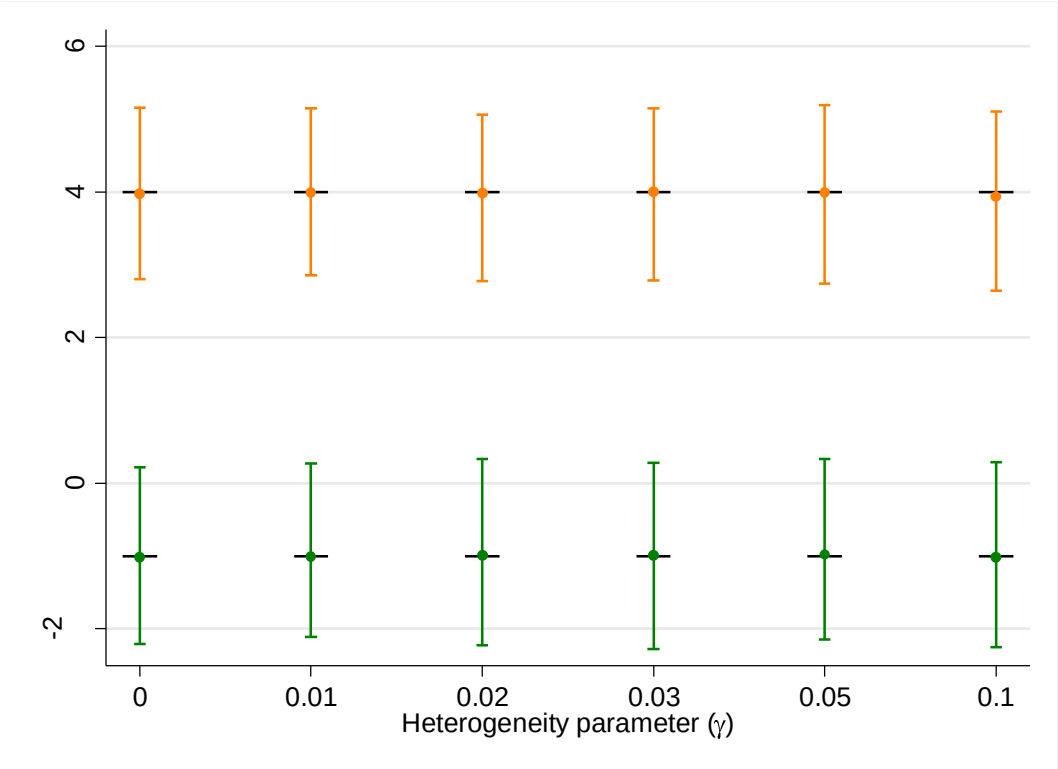

Figure S11. LATE intercepts (green) and LATE slopes (orange) for different degrees of heterogeneity of β_{xg} in unknown subgroups in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

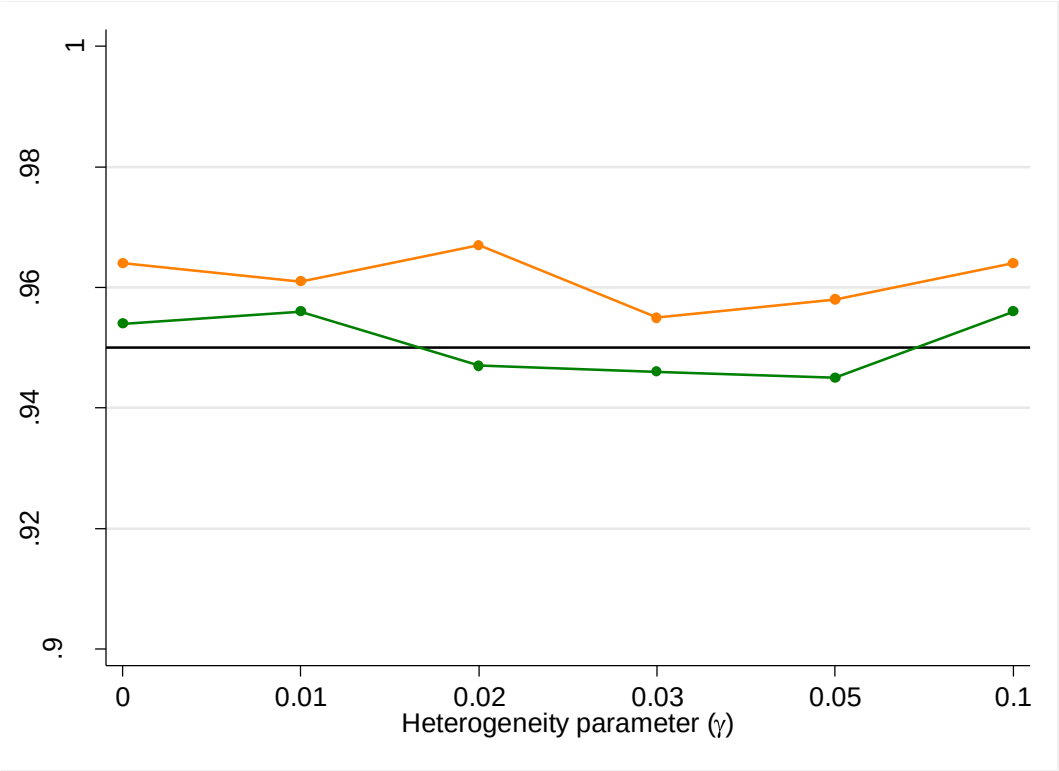

Figure S12. 95% confidence interval coverage for LATE intercepts (green) and LATE slopes (orange) for different degrees of heterogeneity of β_{xg} in unknown subgroups in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Horizontal solid line represents the target value (95%).

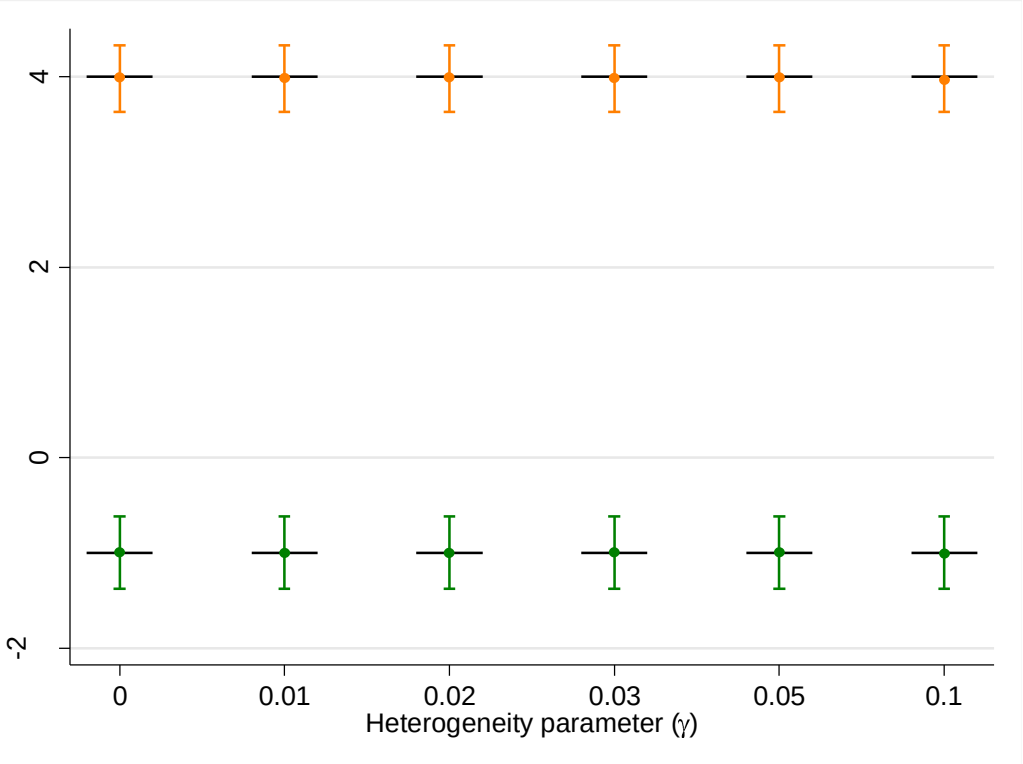

Figure S13. LATE intercepts (green) and LATE slopes (orange) for different degrees of heterogeneity of β_{xg} in known subgroups in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

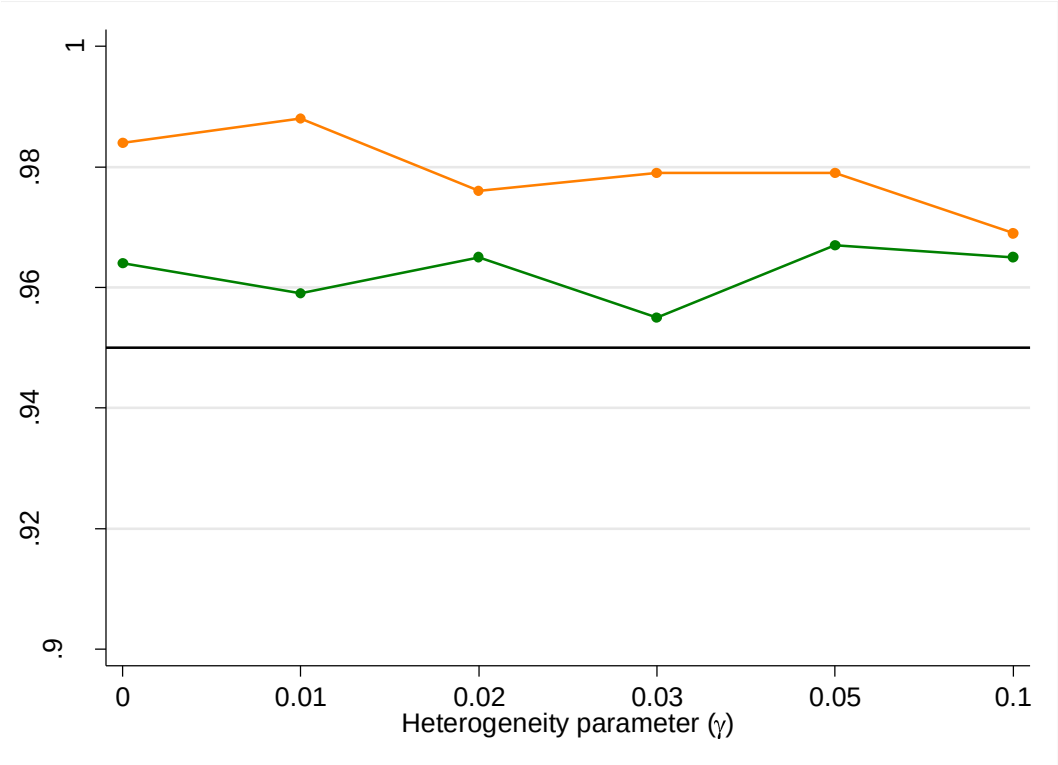

Figure S14. 95% confidence interval coverage for LATE intercepts (green) and LATE slopes (orange) for different degrees of heterogeneity of β_{xg} in known subgroups in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Horizontal solid line represents the target value (95%).

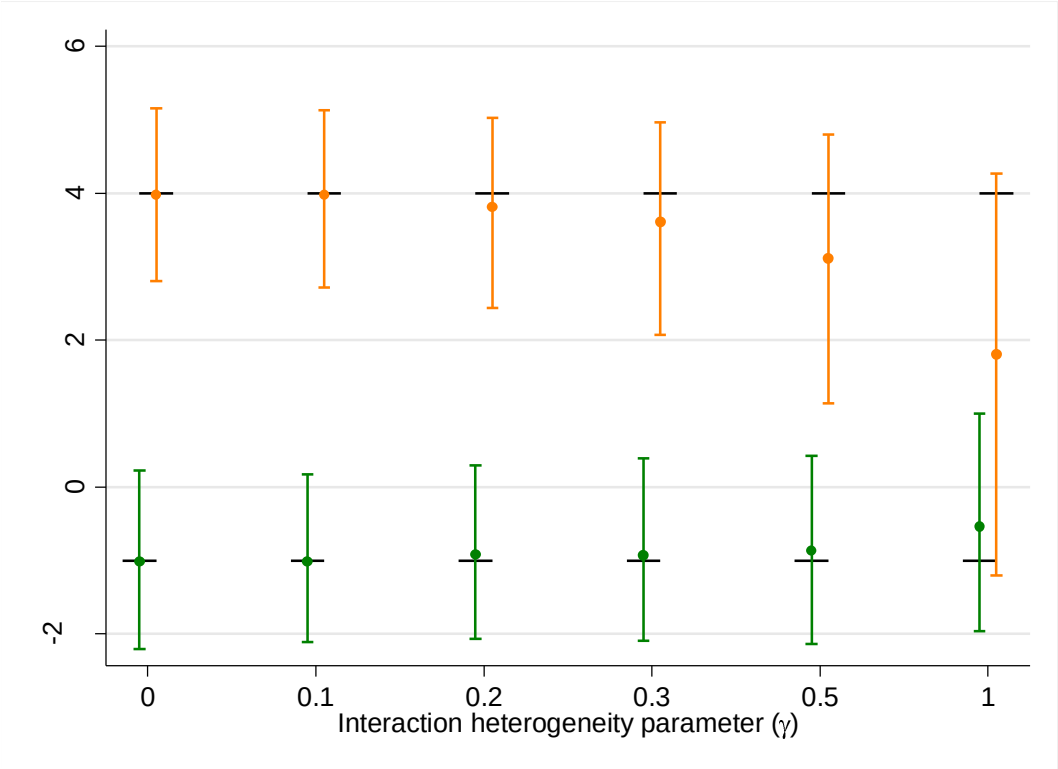

Figure S15. LATE intercepts (green) and LATE slopes (orange) for different degrees of G-U interaction which varies between unknown subgroups in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Points represent means and bars represent 95% of the data. Horizontal solid lines represent target values.

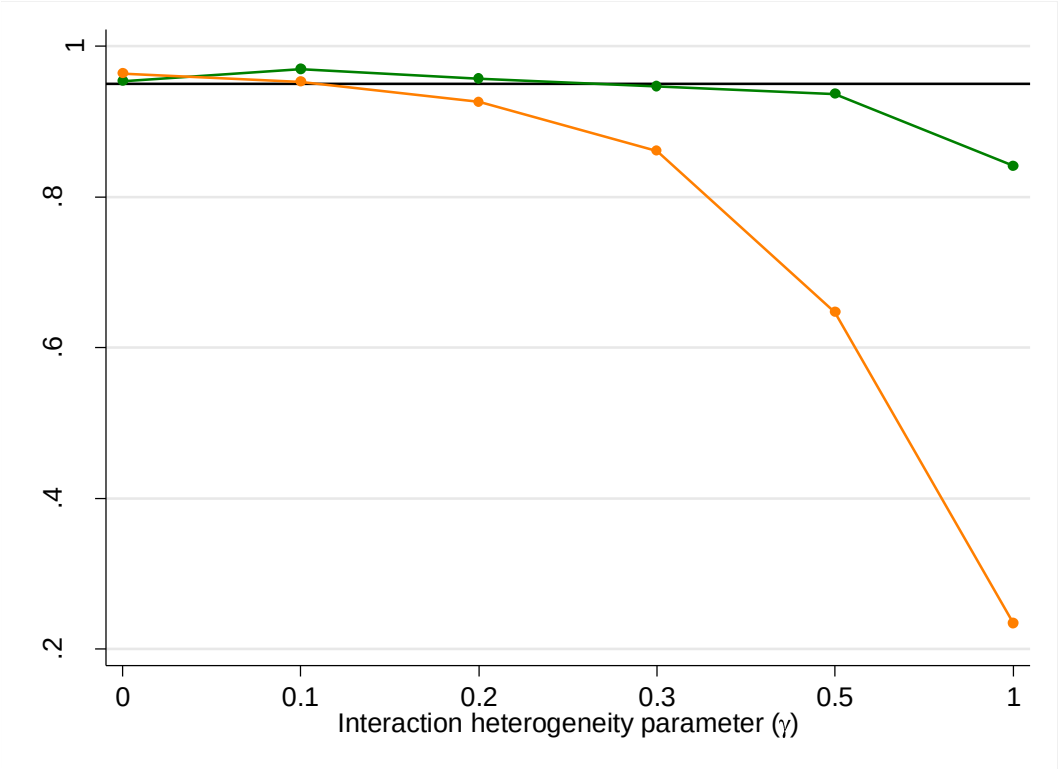

Figure S16. 95% confidence interval coverage for LATE intercepts (green) and LATE slopes (orange) for different degrees of G-U interaction which varies between unknown subgroups in the quadratic data generating model with linear coefficient = -1 and quadratic coefficient = 2. Horizontal solid line represents the target value (95%).

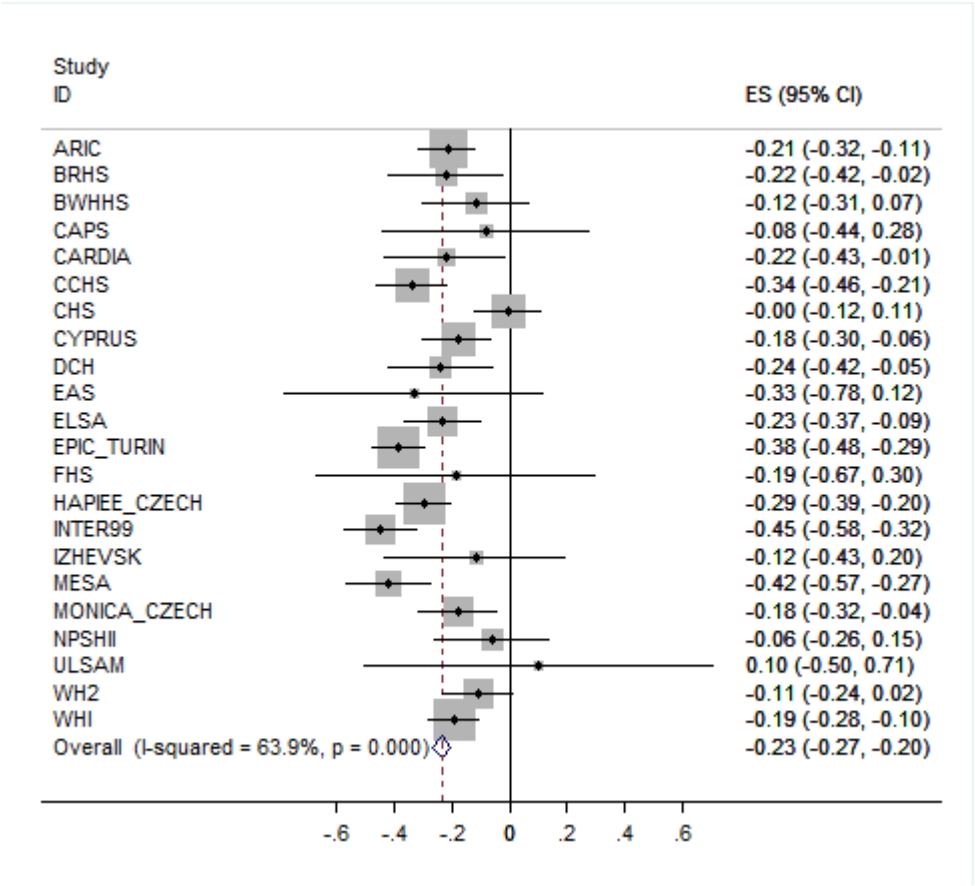

Figure S17. Study-specific estimates of the association between *rs1229984* and log-alcohol.

References

1. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. 2011.
2. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011; **40**: 755-64.