Alcohol consumption and mate choice in UK Biobank: comparing observational and Mendelian randomization estimates

Authors: Laurence J Howe*, [1, 3] Daniel J Lawson, [1] Neil M Davies, [1] Beate St. Pourcain, [2] Sarah J Lewis, [1] George Davey Smith, [1] and Gibran Hemani [1]

[1] Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, United Kingdom

[2] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

[3] Institute of Cardiovascular Science, University College London, United Kingdom

* Correspondence to laurence.howe@ucl.ac.uk

Abstract

Alcohol use is correlated within spouse-pairs, but it is difficult to disentangle the effects of alcohol consumption on mate-selection from social factors or cohabitation leading to spouses becoming more similar over time. We hypothesised that genetic variants related to alcohol consumption may, via their effect on alcohol behaviour, influence mate selection.

Therefore, in a sample of over 47,000 spouse-pairs in the UK Biobank we utilised a well-characterised alcohol related variant, rs1229984 in *ADH1B*, as a genetic proxy for alcohol use. We compared the phenotypic correlation between spouses for self-reported alcohol use with the association between an individual's self-reported alcohol use and their partner's rs1229984 genotype using Mendelian randomization. This was followed up by an exploration of the spousal genotypic concordance for the variant.

We found strong evidence that both an individual's self-reported alcohol consumption and rs1229984 genotype are associated with their partner's self-reported alcohol use. The Mendelian randomization analysis, found that each unit increase in an individual's weekly alcohol consumption increased their partner's alcohol consumption by 0.29 units (95% C.I. 0.20, 0.38; $P=2.15 \times 10^{-9}$). Furthermore, the rs1229984 genotype correlated within spouse-pairs, suggesting that some spousal correlation existed prior to cohabitation. Although the SNP is strongly associated with ancestry, our results suggest that this concordance is unlikely to be explained by population stratification. Overall, our findings suggest that alcohol behaviour directly influences mate selection.

Introduction

Human mate choice is highly non-random; spouse-pairs are generally more phenotypically similar than would be expected by chance [1-6]. Previous studies suggest that alcohol related phenotypes, ranging from consumption to alcohol dependence, are highly correlated within spouse-pairs [7-13]. However, the extent to which the spousal correlation is due to the effect of alcohol behaviour on mate selection (assortative mating) is currently unclear. Indeed, the spousal correlation may be related to assortment on other social and environmental factors (social homogamy) or a consequence of an individual's partner influencing their alcohol behaviour after the individuals have paired up (partner interaction effects) [11-13]. The mechanism explaining spousal concordance for alcohol consumption could have important implications. For example, partner interactions over time explaining the spousal concordance would suggest that public health policy may benefit from focusing on couples rather than individuals to reduce population level alcohol intake. **Figure 1** illustrates the possible explanations for spousal correlation on alcohol consumption.

One biological mechanism that partially explains the phenotypic concordance between spouse-pairs is that they are on average more genetically similar across the genome than non-spouse-pairs [14]. Genotypes implicated in the aetiology of height, education, blood pressure and several chronic diseases have been shown to be correlated within spouse-pairs [15-18]. It is not known whether genetic variants implicated in alcohol metabolism, via their effect on alcohol behaviour, contribute to mate selection.

Alcohol behaviour has been shown to be highly heritable; Genome-wide Association Studies (GWAS) have identified more than 15 loci implicated in either the aetiology of alcohol dependence [19-23] or alcohol consumption volume [21 24-27]. Notably, genetic variants in the Alcohol Dehydrogenase (*ADH*) and Aldehyde Dehydrogenase (*ALDH*) gene families are associated with differences in alcohol consumption. For example, *ADH1B* is involved in the production of enzymes that oxidise alcohol and so individuals with certain alleles may find alcohol consumption unpleasant, resulting in lower intake. Similarly, a genetic variant in *ALDH2*, rare in non-east Asian populations, is associated with a “flush reaction” to alcohol [28 29].

Alcohol consumption-related genetic variants can be useful to determine the most likely explanation for the spousal phenotypic correlation for alcohol use, by analogy with Mendelian randomization studies [30]. Genetic variants for alcohol consumption are in theory less susceptible to confounding from socioeconomic and behavioural factors than measured alcohol consumption so can be used to rule out the possibility that social homogamy is driving the spousal phenotypic correlation [30 31]. The timing of the effects of alcohol consumption can be discerned by evaluating the spousal genotypic correlation for alcohol use-related variants. Genotypic correlation would imply that an effect exists prior to pairing, suggesting that some degree of the spousal phenotypic correlation is attributable to assortative mating (**Figure 2**).

In this study we aimed to explore spousal similarities for alcohol consumption using observational and genetic data. First, we estimated the association of an individual’s self-reported alcohol use with the self-reported alcohol use of their partner. Second, we used a Mendelian randomization framework to estimate the effect of an individual’s alcohol use on their spouse’s alcohol use. Here, we used

their partner's rs1229984 genotype, a missense mutation in *ADH1B* strongly associated with alcohol consumption as an instrumental variable for self-reported alcohol consumption. Third, we estimated the association of rs1229984 genotype between spouses, to evaluate the timing of possible causal effects, and investigate the possibility of bias from population stratification. As a positive control, to demonstrate the validity of derived spouse pairs and the usage of a Mendelian randomization framework, we also analysed height, known to be correlated between spouses, using similar methods.

Materials and Methods

Study participants

UK Biobank

UK Biobank is a large-scale cohort study, including 502,655 participants aged between 40-69 years. Study participants were recruited from 22 recruitment centres across the United Kingdom between 2006 and 2010 [32].

European sub-sample and spouse pairs

Spouse information is not explicitly available, therefore we used similar methods to previous studies [15-17] to identify spouse-pairs in the UK Biobank. Firstly, the data-set was restricted to a subset of 463,827 individuals of recent European descent. Individuals of non-European descent were removed based on a k-means cluster analysis on the first 4 genetic principal components [33].

Next, household sharing information was used to extract pairs of individuals who (a) report living with their spouse, (b) report the same length of time living in the house, (c) report the same number of occupants in the household, (d) report the same number of vehicles, (e) report the same accommodation type and rental

status, (f) have identical home coordinates (rounded to the nearest km), (g) are registered with the same UK Biobank recruitment centre and (h) both have available genotype data. If more than two individuals shared identical information across all variables, these individuals were excluded from analysis. At this stage, we identified 52,471 potential spouse-pairs.

We excluded 4,866 potential couples who were the same sex (9.3% of the sample). To reduce the possibility that identified spouse-pairs are in fact related, non-spouse pairs; we removed 3 pairs reporting the same age of death for both parents. Then we constructed a genetic relationship matrix (GRM) amongst derived pairs and removed 53 pairs with estimated relatedness (IBD >0.1). To construct the GRM; we used a pool of 78,341 markers which were derived by LD pruning (50KB, steps of 5 KB, r2<0.1) 1,440,616 SNPs from the HapMap3 reference panel [34] using the 1000 Genomes CEU genotype data [35] as a reference panel. The final-sample included 47,549 spouse-pairs.

Height

At baseline, the height (cm) of UK Biobank participants was measured using a Seca 202 device at the assessment centre. Measured height was used as a positive control for the application of a Mendelian randomization framework in the context of assortative mating.

Self-reported alcohol variables

At baseline, study participants completed a questionnaire. Participants were asked to describe their current drinking status (never, previous, current, prefer not to say) and estimate their current alcohol intake frequency (daily or almost daily, three

or four times a week, once or twice a week, one to three times a month, special occasions only, never, prefer not to say). Individuals reporting a current intake frequency of at least “once or twice a week” were asked to estimate their average weekly intake of a range of different alcoholic beverages (red wine, white wine, champagne, beer, cider, spirits, fortified wine).

From these variables, we derived three measures: ever or never consumed alcohol (current or former against never), a binary measure of current drinking for self-reported current drinkers (three or more times a week against less than three times a week) and an average intake of alcoholic units per week, derived by combining the self-reported estimated intakes of the different alcoholic beverages consumptions across the five drink types, as in a previous study [24]. The questionnaire used the following measurement units for each of the five alcoholic drink types: measures for spirits, glasses for wines and pints for beer/cider which were estimated to be equivalent to 1, 2 and 2.5 units respectively. Individuals reporting current intake frequency of “one to three times a month”, “special occasions only” or “never” (for whom this phenotype was not collected), were assumed to have a weekly alcohol consumption volume of 0.

Genotyping

488,377 UK Biobank study participants were assayed using two similar genotyping arrays, the UK BiLEVE Axiom™ Array by Affymetrix1 (N= 49,950) and the closely-related UK Biobank Axiom™ Array (N= 438,427). Directly genotyped variants were pre-phased using SHAPEIT3 [36] and then imputed using Impute4 using the UK10K [37], Haplotype Reference Consortium [38] and 1000 Genomes Phase 3 [35] reference panels. Post-imputation, data were available on approximately ~96 million genetic variants [39 40].

Statistical analysis

Phenotypic spousal correlation for height

To verify the validity of the derived spouse-pair sample, we evaluated the spousal phenotypic correlation for height. Previous studies have found strong evidence of spousal correlation for height, so comparable results would be consistent with derived spouses being genuine. The spousal phenotypic correlation was estimated using a linear regression of an individual's height against the height of their partner, adjusting for sex. With one unique phenotype pairing within couples, each individual in the data-set was included only once as either the reference individual or their partner.

Phenotypic spousal correlation for self-reported alcohol behaviour

To evaluate the phenotypic correlation on alcohol use we compared self-reported alcohol behaviour between spouses. We estimated the spousal correlation for the two binary measures (ever or never consumed alcohol, three or more times a week) using a logistic regression of the relevant variable for an individual against the relevant variable for their partner, adjusting for sex. Similarly, linear regression was used to estimate the spousal-correlation for continuous weekly alcohol consumption volume, adjusting for sex. Spouse-pairs with any missing phenotype data, or where one or more spouses reported their weekly alcohol consumption volume to be more than five standard deviations away from the mean (calculated using the sample of individuals with non-zero weekly drinking) were removed from relevant analyses. With one unique phenotype pairing within couples, each individual in the data-set was included only once as either the reference individual or their partner.

Mendelian randomization: Genetically influenced height and measured height of partner

We validated the application of an Mendelian randomization approach to assortative mating using height as a positive control; genotypes influencing height have previously demonstrated to be highly correlated between spouse-pairs [15]. As a measure of genetically influenced height, we started with 382 independent SNPs, generated using LD clumping (r^2<0.001) in MR-Base [41], from a recent Genome-wide Association Study (GWAS) of adult height in Europeans [42].

For the purposes of the Mendelian randomization analysis, we restricted analyses to spouse-pairs with complete measured height data and genotype data. First, we estimated the association between the 382 SNPs and height in the same individual, using the spouse-pair sample with sex included as a covariate. We removed 23 SNPs that were not strongly associated with height (P> 0.05) or with inconsistent directions of effect between our sample and the GWAS summary statistics. Second, we estimated the association between the 359 remaining SNPs and spousal height. PLINK [43] was used to estimate the SNP-phenotype associations also including sex as a covariate. We then estimated the effect of a 1 cm increase in an individual's height on their partner's height using the TwoSampleMR R package [41] and the internally derived weights described above. The fixed-effects Inverse-Variance Weighted (IVW) method was used as the primary analysis. Cochran's Q test and the I^2 statistic were used to test for heterogeneity in the fixed-effects IVW [44]. MR Egger [45] was used to test for directional pleiotropy. The weighted median [46] and mode [47] were used to test the consistency of the effect estimate. With two unique pairings between genotype and phenotype in each couple, each individual in the data-set was included twice as both the reference individual and as the partner.

Mendelian randomization: Genetically influenced alcohol consumption volume and self-reported alcohol consumption of partner

We then applied the Mendelian randomization framework to investigate if an individual's genotype at rs1229984 in *ADH1B* affects the self-reported alcohol consumption volume of their partner. Given the rarity of individuals homozygous for the minor allele in European populations, the MAF is 2.9% in the 1000 Genomes CEU population [35], we assumed a dominant model consistent with previous studies [48 49]. We restricted analysis to spouse-pairs where both members had genotype data, and one or more members had self-reported alcohol consumption volume. First, we estimated the association of the rs1229984 genotype with alcohol consumption in the same individual after adjusting for sex. Second, we estimated the association between rs1229984 and spousal alcohol consumption after adjusting for sex. PLINK [43] was used to estimate the SNP-phenotype associations. We then estimated the effect of a 1 unit increase in an individual's weekly alcohol consumption volume on the same variable in their partner. The Wald ratio estimate was obtained using mr_wald_ratio function in the TwoSample MR R package [41] using internally derived weights. Sensitivity analyses were limited due to the use of a single genetic instrument. With two unique pairings between genotype and phenotype in each couple, each individual in the data-set was included twice as both the reference individual and as the partner.

Spousal genotypic correlation for rs1229984 genotype

We then investigated properties of the rs122984 variant in the UK Biobank that may be relevant to assortative mating. Starting with the UK Biobank subset of 463,827 individuals of recent European descent, we removed 78,540 related

individuals (relevant methodology has been described previously [33]) and tested Hardy-Weinberg Equilibrium (HWE) in the resulting sample of 385,287 individuals. We then investigated the association of the SNP with genetic principal components and birth coordinates. As a sensitivity analysis we also restricted the sample to a more homogeneous sample of white British individuals, provided by the UK Biobank, and repeated analyses. With one unique genotype pairing within couples, each individual in the data-set was included only once as either the reference individual or their partner.

We then estimated the genotypic concordance between derived spouse-pairs for rs1229984 genotype using logistic regression, again assuming a dominant model. As a sensitivity analysis, we then investigated the possibility that spousal-concordance for rs1229984 was driven by fine-scale assortative mating due to geography, which is itself associated with genetic variation within the UK [50 51]. For this, we restricted the sample to include only 28,693 spouse-pairs born within 100 miles of each other. The spouse-pairs were then stratified into the 22 different UK Biobank recruitment centres and logistic regression analyses were re-run to estimate the spousal-concordance of the *ADH1B* genotype by centre. Geographical patterns of heterogeneity across the different UK Biobank recruitment centres would provide evidence of population stratification.

A list of derived spouse-pairs has been returned to UK Biobank

Results

Phenotypic spousal correlation for height

Measured height was strongly correlated between spouse-pairs. In a sample of 47,377 spouse-pairs, a 1 unit increase in an individual's height was associated with a 0.24 unit increase (95% C.I. 0.23, 0.25, $P<10^{-16}$) in their partner's height. This result is consistent with previous findings [52], validating the derived spouse pairs.

Phenotypic spousal correlation for self-reported alcohol behaviour

The majority of derived spouse-pairs had complete data for relevant self-reported alcohol behaviour phenotypes. Strong evidence was found for phenotypic correlation between spouse-pairs for all self-reported alcohol variables. Amongst 47,510 spouse-pairs, an individual self-reporting as a never-drinker was associated with increased odds (OR 14.06, 95% C.I., 11.95, 16.50 $P<10^{-16}$) of their partner self-reporting as a never-drinker. Similarly, when restricting to 42,844 pairs who both reported being current-drinkers, an individual drinking three or more times a week had increased odds (OR 6.64, 95% C.I., 6.34, 6.94 $P<10^{-16}$) of their partner also drinking three or more times a week.

For self-reported alcohol consumption volume; 47,510 spouse-pairs had either complete phenotype data or reported their consumption frequency as less than weekly (in which case their weekly volume was assumed to be 0). After removing 189 pairs with outlying values from one or more members, the final sample included 47,321 spouse-pairs. In this sample, each unit increase in an individual's weekly alcohol consumption volume was associated with a 0.38-unit increase (95% C.I. 0.37, 0.38 $P<10^{-16}$) in the same variable in their partner.

Genetically influenced height and height of partner

The application of Mendelian randomization to spousal height was consistent with the previous evidence for assortative mating on height. Across 47,377 spouse-pairs, a 1 cm increase in an individual's height was associated with a 0.19 cm increase in their partner's height (95% C.I. 0.18, 0.21; $P=7.0x10^{-114}$). The I^2 statistic (2.9%) and Cochran's Q test (P=0.64) suggested consistent effects across SNPs, and estimates were consistent across the weighted median, weighted modal and MR-Egger estimators with the MR-Egger intercept test finding no strong evidence for directional pleiotropy (**Table 1)**.

Table 1: Mendelian randomization estimates for the effect of a 1 cm increase in height on partner's height

Test	**Interpretation**	**Estimate (95% C.I.)**	**P-value**
Inverse variance weighted	Primary causal estimate[1]	0.19 (0.18, 0.21)	$7.0x10^{-114}$
Heterogeneity of Inverse variance weighted	Balanced pleiotropy	I^2=2.9%	0.64
MR-Egger	Intercept test for directional pleiotropy[2]	0.005 (-0.003, 0.012)	0.24
	Regression estimate[1]	0.17 (0.13, 0.21)	$1.1x10^{-16}$
Weighted median	Consistency[1]	0.18 (0.15, 0.21)	$3.4x10^{-39}$
Weighted mode	Consistency[1]	0.18 (0.13, 0.23)	$1.1x10^{-10}$

1 Units: mm change in philtrum width per 1-unit log odd increase in liability to nsCL/P

2 Units: Average pleiotropic effect of a nsCL/P genetic variant on philtrum width

Mendelian randomization framework: Genetically influenced alcohol consumption and self-reported alcohol behaviour of partner

To evaluate the degree to which an individual's alcohol consumption is affected by their partner's genetically influenced alcohol consumption, we used the same sample of 47,321 spouse-pairs from the previous phenotypic correlation analysis. In this sample, individuals with two copies of the *ADH1B* major allele consumed 4.95 more alcoholic units a week (95% C.I. 4.48, 5.42; $P<10^{-16}$) than the reference group (individuals with one or no copies). The partners of individuals with two *ADH1B* major alleles consumed 1.44 more units a week (95% C.I. 0.97, 1.91; $P=3.61x10^{-10}$) than the reference group. After scaling the estimate using a Wald estimator; a 1 unit increase in an individual's alcohol consumption led to having partner's with alcohol consumption 0.29 units higher than baseline (95% C.I. 0.20, 0.38; $P=2.15x10^{-9}$). This effect is slightly lower than the phenotypic estimate of 0.38 units (95% C.I. 0.37, 0.38) although confidence intervals overlap.

Spousal genotypic correlation for rs1229984 genotype

Characteristics of rs1229984 in the UK Biobank

In the sample of 385,287 individuals of recent European descent, the MAF of rs1229984 was 2.8% and very strong evidence was found for the SNP violating HWE (Chi^2 = 275, $P <10^{-16}$) due to fewer heterozygotes compared to expectation (expected=20,972, observed=20,194). However, when restricting to the sample of 337,114 individuals of white British descent, the MAF of rs1229984 was 2.2% and there was less clear evidence for the SNP violating HWE (Chi^2 = 2.0, P=0.16) and there were actually more heterozygotes compared to expected (expected= 14,506 observed=14,743) (**Supplementary Table 1**). Evidence was found of allele

frequency differences for rs1229984 between the two samples (Chi^2=445, $P<10^{-16}$) suggesting that population substructure differences may explain the HWE results.

The SNP was found to be strongly associated with both genetic principal components and birth coordinates in both samples. In the less restrictive European sample, individuals with 1 or more copies of the minor allele of rs1229984 on average were born 25.2 miles farther south (95% C.I. 22.7, 27.7) and 13.6 miles farther east (95% C.I. 12.4, 14.9) than individuals with no copies of the minor allele. The SNP was similarly associated with principal components and birth coordinates in the sample of white British descent although there were differences in effect estimates between the two samples (**Supplementary Table 2**).

Genetic correlation analysis

Amongst 47,549 spouse-pairs, strong concordance was observed for the genotype of rs1229984. Individuals with no copies of the minor allele had increased odds of having a partner with no copies of the minor allele (OR 1.30; 95% C.I. 1.11, 1.53; P=0.00129).

As a sensitivity analysis, we restricted the sample to 28,693 spouse-pairs born within 100 miles of each other and stratified spouse-pairs by the 22 different UK Biobank recruitment centres. Of the 22 centres, 5 centres were omitted from the meta-analysis because the limited sample sizes led to convergence issues in regression. A fixed-effects meta-analysis was then used to estimate the spousal-concordance across the remaining 17 centres and 27,831 spouse-pairs. Evidence was found of spousal concordance for rs1229984 (OR 1.34; 95% C.I. 1.07,1.68; P=0.0123), consistent with the previous analysis. Cochran's Q test for heterogeneity

across the logOR suggested no strong evidence for heterogeneity (P= 0.88) across the different centres (**Table 2**).

Table 2: Meta-analysis of spousal-concordance for rs1229984 across the UK Biobank recruitment centres

Recruitment Centre	**Number of spouse-pairs born within 100 miles of each other**	**OR (95% C.I.)**
Stockport	15	N/A[1]
Manchester	663	1.00 (0.23, 4.33)
Oxford	670	0.86 (0.20, 3.72)
Cardiff	933	1.52 (0.45, 5.12)
Glasgow	1048	2.99 (0.87, 10.3)
Edinburgh	611	N/A[1]
Stoke	1217	0.77 (0.18, 3.24)
Reading	1352	1.09 (0.43, 2.78)
Bury	2253	1.25 (0.57, 2.77)
Newcastle	2981	0.41 (0.10, 1.68)
Leeds	2565	1.74 (0.89, 3.41)
Bristol	2119	1.20 (0.54, 2.63)
St Bartholomew's Hospital	122	N/A[1]
Nottingham	2344	1.38 (0.65, 2.89)
Sheffield	2262	1.84 (0.87, 3.90)
Liverpool	2633	1.14 (0.49, 2.66)
Middlesbrough	1477	1.04 (0.32, 3.39)
Hounslow	838	1.87 (0.85, 4.11)
Croydon	1035	1.69 (0.70, 4.09)
Birmingham	1441	0.66 (0.20, 2.15)
Swansea	85	N/A[1]
Wrexham	29	N/A[1]
Combined (Fixed effects)	**27,831**	**1.34 (1.07, 1.68) P=0.0123**

1 Logistic regression estimates did not converge due to limited sample sizes, these studies were excluded from the meta-analysis.

Discussion

In this study, we used a large sample of derived spouse-pairs in a UK-based cohort to demonstrate that an individual's self-reported alcohol use and their genotype for an alcohol implicated variant, rs1229984 in *ADH1B*, are associated with their partner's self-reported alcohol use. Furthermore, we showed that the genotype of a variant influencing alcohol metabolism, rs1229984, is correlated within spouse-pairs. There are three possible explanations for our findings. First, that rs1229984 influences alcohol behaviour, which has a downstream effect on mate selection. Second, that a participant's alcohol use is influenced by their partner's alcohol use. Third, that given the strong association of the SNP with both genetic principal components and birth coordinates, the spousal concordance is related to factors influencing social homogamy, independent of alcohol behaviour, such as place of birth, ancestry or socio-economic status. Indeed, the allele frequency of rs1229984 was found to deviate between European and white British subsets of the UK Biobank.

However, we presented evidence suggesting that a substantial proportion of the spousal concordance is likely to be explained by the biological effects of the variant on alcohol consumption. Firstly, we have tested the association between a causal SNP for alcohol consumption, and not the measured consumption itself, thereby avoiding any post-birth confounding factors suggesting that alcohol use has a direct effect on spousal alcohol use. Secondly, because rs1229984 is correlated between spouses, there must be some degree of assortment on alcohol consumption prior to cohabitation. This suggests that the spousal correlation cannot be entirely due to the effect of the individual's alcohol consumption behaviour on their spouse's behaviour. Thirdly, in a sensitivity analysis, we controlled for shared

ancestry, which could have induced confounding, by excluding spouse-pairs born more than 100 miles apart, and the within sub-population effect estimates remained consistent.

The strong evidence for spousal-correlation on the variant has implications for conventional Mendelian randomization studies (i.e. estimating the causal effect of an exposure on an outcome) [30] which use the SNP as a genetic proxy for alcohol intake [48]. Assortative mating could lead to a violation of the Mendelian randomization assumption, that the genetic instrument for the exposure is not strongly associated with confounders of the exposure-outcome relationship. If both genetic and environmental factors affect alcohol consumption, then assortative mating on alcohol consumption would induce associations between genetic and environmental factors in the offspring, with the strength of association dependent on the degree of assortative mating [53].

Interestingly, the minor allele of rs1229984 (i.e. associated with lower alcohol consumption) has been previously found to be positively associated with years in education [48] and socio-economic related variables, such as the Townsend deprivation index and number of vehicles in household [54 55]. These associations may be down-stream causal effects of alcohol consumption, which implies that some of the spousal concordance for alcohol consumption could be explained by assortative mating on educational attainment [15] or alternatively these associations may reflect maternal genotype and intrauterine effects [56]. Over time, assortative mating on alcohol consumption may further strengthen the associations between rs1229984 and socio-economic related variables [53]. Of further interest is that the variant has

previously been shown to be under selection [57] suggesting that the variant has historically had a substantial effect on reproductive fitness.

The analyses in this study extended previous work on the correlation between spouse-pairs for alcohol behaviour [7-12] by comparing the phenotypic correlation with analyses utilising a genetic variant strongly associated with alcohol consumption. A major strength of this study is the use of three distinct methods with different non-overlapping limitations, allowing for improved inference by triangulating the results from the different methods [58]. First, we evaluated the spousal phenotypic correlation for self-reported alcohol consumption, second we investigated the effect of an individual's rs1229984 genotype on the alcohol consumption of their spouse using Mendelian randomization and third we demonstrated spousal genotypic correlation for rs1229984. The use of the UK Biobank data-set was a considerable strength for these analyses because of the low frequency of the rs1229984 minor allele; the large scale of the UK Biobank allowed for the identification of thousands of genotyped spouse-pairs. A further strength of these analyses is that we have demonstrated the utility of a Mendelian randomization framework for application to assortative mating by applying it to height and alcohol use. A similar approach using polygenic risk scores has previously demonstrated assortative mating on educational attainment [18]. However, the use of Mendelian randomization has a notable advantage over polygenic approaches because of the possibility of using various sensitivity analyses to test for heterogeneity and consistency of the effect estimate [45-47].

There are several limitations of this study. First, although spouse-pairs were identified using similar methods to previous studies [15-17], the identified spouse-pairs have not been confirmed. However, the phenotypic spousal correlation estimate for

height found in this study is highly concordant with previous estimates [52], consistent with derived couples being genuine. Second, despite follow-up analyses, it is difficult to definitively prove that the spousal concordance is a direct result of assortative mating on alcohol consumption. Assortment independent of alcohol use cannot be completely ruled out and down-stream pleiotropic effects of the variant may influence mate selection. Third, the use of a single genetic instrument in the Mendelian randomization analysis, limited the use of sensitivity analyses [45-47] and meant it is not possible to infer similar associations for other alcohol-implicated variants. Finally, it is difficult to extrapolate the results of this study in the UK Biobank to non-European populations. This is because of potential contextual influences; for example, in East Asian populations, males are much more likely to consume alcohol than females [59] [60]. Additionally, there is some evidence that the effect of genetic contributors to alcohol varies across different populations [27].

To conclude, our results suggest that there is non-random mating on rs1229984 in *ADH1B*, likely related to the effect of the variant on alcohol behaviour. These results suggest that alcohol use influences mate selection and argue for a more nuanced approach to considering social and cultural factors when examining causality in epidemiological studies. Further research investigating other alcohol-implicated variants, and other societies and ethnicities, would strengthen these conclusions.

Figure titles and descriptions

Figure 1 Possible explanations for spousal correlation on alcohol use.

(A) Assortative mating. Alcohol behaviour influences mate selection; individuals are more likely to select a mate with similar alcohol consumption.

(B) Social homogamy or confounding. An unknown confounder influence mate selection independent of alcohol behaviour. For example, ancestry or socio-economic status may influence both alcohol use and mate choice.

(C) Partner interaction effects. As spouse-pairs co-habitate their alcohol behaviour becomes more similar over time.

Figure 2 Phenotypic correlation, Mendelian randomization and genotypic correlation

(1) Phenotypic correlation. Spousal correlation for alcohol use could be explained by any of the three previously mentioned possibilities: assortative mating, social homogamy or partner interaction effects.

(2) Mendelian randomization framework. An association between an individual's alcohol influencing genotype and their spouse's alcohol use would suggest that the spousal correlation is explained by either assortative mating or partner interaction effects. Genetic variants are unlikely to be associated with socio-economic confounders suggesting that social homogamy is unlikely.

(3) Genotypic correlation. Genotypic correlation for alcohol related genetic variants would suggest that some degree of the spousal correlation is explained by assortative mating. Partner interaction effects cannot lead to genotypic correlation because genotypes are fixed from birth.

Acknowledgements

LJH was a Medical Research Council funded PhD student at the University of Bristol and is now funded by the British Heart Foundation and University College London. NMD, SJL and GDS work in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/1) which is supported by the Medical Research Council and the University of Bristol. DJL [WT104125MA] and GH [208806/|/17/|] are both supported by the Wellcome Trust. UK Biobank data access was granted through the MR base application 15825 (PI: Dr Philip Haycock).

References

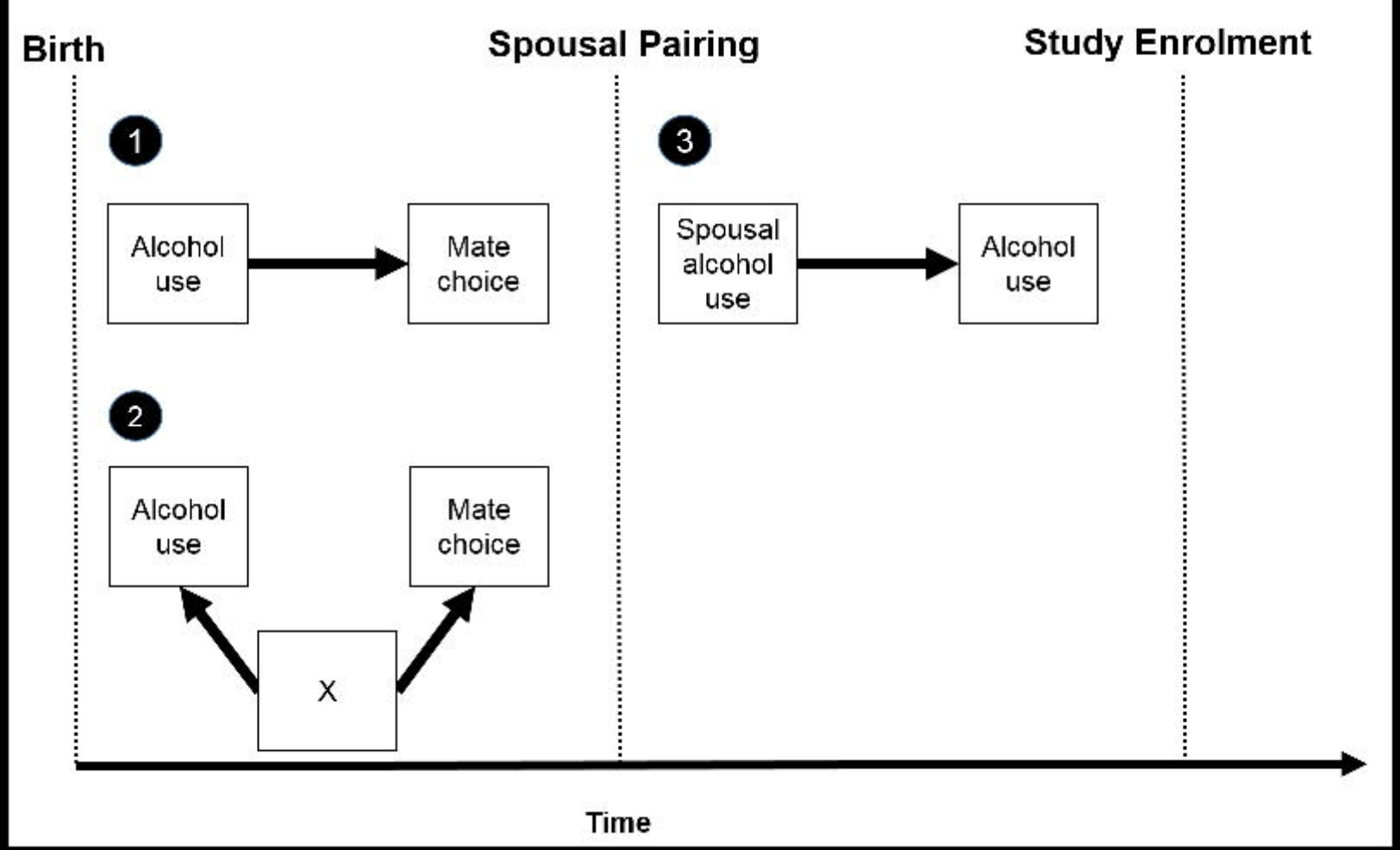
1. Vandenberg SG. Assortative mating, or who marries whom? Behavior Genetics 1972;**2**(2-3):127-57.
2. Buss DM. Human mate selection: Opposites are sometimes said to attract, but in fact we are likely to marry someone who is similar to us in almost every variable. American Scientist 1985;**73**(1):47-51.
3. Mare RD. Five decades of educational assortative mating. American Sociological Review 1991:15-32.
4. Silventoinen K, Kaprio J, Lahelma E, et al. Assortative mating by body height and BMI: Finnish twins and their spouses. American Journal of Human Biology 2003;**15**(5):620-27.

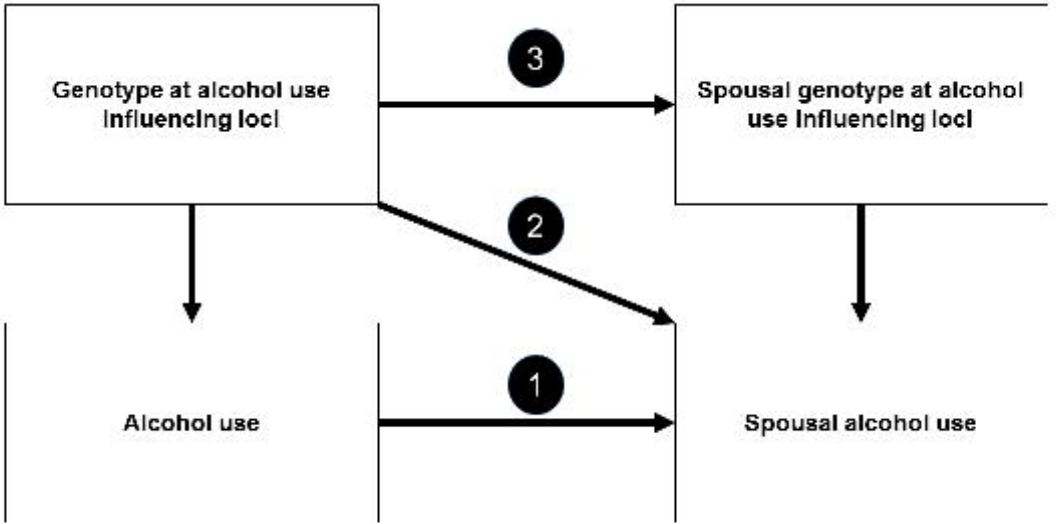
5. Krueger RF, Moffitt TE, Caspi A, et al. Assortative mating for antisocial behavior: Developmental and methodological implications. Behavior Genetics 1998;**28**(3):173-86.
6. Watson D, Klohnen EC, Casillas A, et al. Match makers and deal breakers: Analyses of assortative mating in newlywed couples. Journal of Personality 2004;**72**(5):1029-68.
7. Agrawal A, Heath AC, Grant JD, et al. Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. Behavior Genetics 2006;**36**(4):553-66.
8. Hall RL, Hesselbrock VM, Stabenau JR. Familial distribution of alcohol use: II. Assortative mating of alcoholic probands. Behavior Genetics 1983;**13**(4):373-82.
9. Hall RL, Hesselbrock VM, Stabenau JR. Familial distribution of alcohol use: I. Assortative mating in the parents of alcoholics. Behavior Genetics 1983;**13**(4):361-72.
10. McLeod JD. Spouse concordance for alcohol dependence and heavy drinking: Evidence from a community sample. Alcoholism: Clinical and Experimental Research 1993;**17**(6):1146-55.
11. Reynolds CA, Barlow T, Pedersen NL. Alcohol, tobacco and caffeine use: spouse similarity processes. Behavior Genetics 2006;**36**(2):201.
12. Grant JD, Heath AC, Bucholz KK, et al. Spousal concordance for alcohol dependence: evidence for assortative mating or spousal interaction effects? Alcoholism: Clinical and Experimental Research 2007;**31**(5):717-28.
13. Ask H, Rognmo K, Torvik FA, et al. Non-random mating and convergence over time for alcohol consumption, smoking, and exercise: the Nord-Trøndelag Health Study. Behavior Genetics 2012;**42**(3):354-65.
14. Domingue BW, Fletcher J, Conley D, et al. Genetic and educational assortative mating among US adults. Proceedings of the National Academy of Sciences 2014;**111**(22):7996-8000.
15. Robinson MR, Kleinman A, Graff M, et al. Genetic evidence of assortative mating in humans. Nature Human Behaviour 2017;**1**:0016.
16. Rawlik K, Canela-Xandri O, Tenesa A. Indirect assortative mating for human disease and longevity. bioRxiv 2017:185207.
17. Tenesa A, Rawlik K, Navarro P, et al. Genetic determination of height-mediated mate choice. Genome Biology 2016;**16**(1):269.
18. Hugh-Jones D, Verweij KJ, Pourcain BS, et al. Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores. Intelligence 2016;**59**:103-08.
19. Gelernter J, Kranzler H, Sherva R, et al. Genome-wide association study of alcohol dependence: significant findings in African-and European-Americans including novel risk loci. Molecular Psychiatry 2014;**19**(1):41.
20. Park BL, Kim JW, Cheong HS, et al. Extended genetic effects of ADH cluster genes on the risk of alcohol dependence: from GWAS to replication. Human Genetics 2013;**132**(6):657-68.
21. Bierut LJ, Goate AM, Breslau N, et al. ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Molecular Psychiatry 2012;**17**(4):445.
22. Bierut LJ, Agrawal A, Bucholz KK, et al. A genome-wide association study of alcohol dependence. Proceedings of the National Academy of Sciences 2010;**107**(11):5082-87.

23. Treutlein J, Cichon S, Ridinger M, et al. Genome-wide association study of alcohol dependence. Archives of General Psychiatry 2009;**66**(7):773-84.
24. Clarke T-K, Adams MJ, Davies G, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N= 112 117). Molecular Psychiatry 2017;**22**(10):1376.
25. Schumann G, Coin LJ, Lourdusamy A, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. Proceedings of the National Academy of Sciences 2011;**108**(17):7119-24.
26. Schumann G, Liu C, O'Reilly P, et al. KLB is associated with alcohol drinking, and its gene product β-Klotho is necessary for FGF21 regulation of alcohol preference. Proceedings of the National Academy of Sciences 2016;**113**(50):14372-77.
27. Jorgenson E, Thai KK, Hoffmann TJ, et al. Genetic contributors to variation in alcohol consumption vary by race/ethnicity in a large multi-ethnic genome-wide association study. Molecular Psychiatry 2017;**22**(9):1359.
28. Thomasson HR, Edenberg HJ, Crabb DW, et al. Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. American Journal of Human Genetics 1991;**48**(4):677.
29. Luczak SE, Glatt SJ, Wall TJ. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in Asians: American Psychological Association, 2006.
30. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology 2003;**32**(1):1-22.
31. Davey Smith G, Lawlor DA, Harbord R, et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. PLoS Medicine 2007;**4**(12):e352.
32. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Medicine 2015;**12**(3):e1001779.
33. Mitchell RE, Hemani G, Dudding T, et al. UK Biobank Genetic Data: MRC-IEU Quality Control, version 1, 13/11/2017. Secondary UK Biobank Genetic Data: MRC-IEU Quality Control, version 1, 13/11/2017 2017.
34. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature 2010;**467**(7311):52-58.
35. Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;**526**(7571):68-74.
36. O'Connell J, Sharp K, Shrine N, et al. Haplotype estimation for biobank-scale data sets. Nature Genetics 2016;**48**(7):817-20.
37. Consortium UK. The UK10K project identifies rare variants in health and disease. Nature 2015;**526**(7571):82-90.
38. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics 2016;**48**(10):1279.
39. Allen NE, Sudlow C, Peakman T, et al. UK biobank data: come and get it: American Association for the Advancement of Science, 2014.
40. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on~ 500,000 UK Biobank participants. bioRxiv 2017:166298.
41. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife 2018;**7**:e34408.

42. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature Genetics 2014;**46**(11):1173.
43. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 2007;**81**(3):559-75.
44. Greco M, Del F, Minelli C, et al. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. Statistics in Medicine 2015;**34**(21):2926-40.
45. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology 2015;**44**(2):512-25.
46. Bowden J, Davey Smith G, Haycock PC, et al. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic Epidemiology 2016;**40**(4):304-14.
47. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International Journal of Epidemiology 2017;**46**(6):1985-98.
48. 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.
49. Zuccolo L, Lewis SJ, Davey Smith G, et al. Prenatal alcohol exposure and offspring cognition and school performance. A 'Mendelian randomization'natural experiment. International Journal of Epidemiology 2013;**42**(5):1358-70.
50. Leslie S, Winney B, Hellenthal G, et al. The fine-scale genetic structure of the British population. Nature 2015;**519**(7543):309.
51. Haworth S, Mitchell R, Corbin L, et al. Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications. bioRxiv 2018:294876.
52. Price RA, Vandenberg SG. Spouse similarity in American and Swedish couples. Behavior Genetics 1980;**10**(1):59-71.
53. Hartwig FP, Davies NM, Davey Smith G. Bias in Mendelian randomization due to assortative mating. Genetic epidemiology 2018.
54. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. bioRxiv 2017:176834.
55. Howe LJ, Sharp GC, Hemani G, et al. Prenatal alcohol exposure and facial morphology in a UK cohort. bioRxiv 2018:302364.
56. Hemani G, Bowden J, Haycock PC, et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. bioRxiv 2017:173682.
57. Galinsky KJ, Bhatia G, Loh P-R, et al. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. The American Journal of Human Genetics 2016;**98**(3):456-72.
58. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. International Journal of Epidemiology 2016;**45**(6):1866-86.
59. Chen L, Davey Smith G, Harbord RM, et al. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. PLoS Medicine 2008;**5**(3):e52.

60. Spiller W, Slichter D, Bowden J, et al. Detecting and correcting for bias in Mendelian randomization analyses using gene-by-environment interactions. bioRxiv 2017:187849.


Birth
Spousal Pairing
Study Enrolment
1
Alcohol use
Mate choice
2
Alcohol use
Mate choice
X
3
Spousal alcohol use
Alcohol use
Time


Genotype at alcohol use influencing loci
3
Spousal genotype at alcohol use influencing loci
2
Alcohol use
1
Spousal alcohol use