Manuscript title: **Evaluating the role of alcohol consumption in breast and ovarian cancer susceptibility using population based cohort studies and Mendelian randomization analyses**

Contribution statements. Please drop your funding details here.

Author contributions

Stuart MacGregor, Eske M. Derks, Penelope M. Webb, Georgia Chenevix-Trench, Stig E. Bojesen and Per Hall designed the study and obtained funding. Mikael Eriksson provided data for the KARMA study cohort. Stig E. Bojesen provided data for the Copenhagen City Heart Study and Copenhagen General Population Study. Jue-Sheng Ong, Jiyuan An, Eske M. Derks, Mikael Eriksson, Stig. E Bojesen and Stuart Macgregor analysed the data. Jue-Sheng Ong, Stuart MacGregor, Stig E. Bojesen and Georgia Chenevix-Trench wrote the first draft of the paper. All other authors contributed to the final version of the paper.

Funding

This work was conducted using the UK Biobank Resource (application number 25331). We thank Scott Wood and John Pearson from QIMR Berghofer for IT support. This work is supported by a project grant (1123248) from the Australian National Health and Medical Research Council (NHMRC). Jue-Sheng Ong received scholarship support from the University of Queensland and QIMR Berghofer Medical Research Institute. Stuart MacGregor, Penelope M. Webb and Georgia Chenevix-Trench are supported by fellowships from the NHMRC. Funding for the CCHS and CGPS were obtained from the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council and Herlev Hospital.

Acknowledgement

We would like to thank the staff and participants of the CCHS, CGPS, KARMA and UK Biobank cohort study. The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The National Institutes of Health (U19 CA148065,

X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and The European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al (2017). The Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PPD/RPCI.07). The scientific development and funding for this project were in part supported by the US National Cancer Institute GAME-ON Post-GWAS Initiative (U19-CA148112). This study made use of data generated by the Wellcome Trust Case Control consortium that was funded by the Wellcome Trust under award 076113. The results published here are in part based upon data generated by The Cancer Genome Atlas Pilot Project established by the National Cancer Institute and National Human Genome Research Institute (dbGap accession number phs000178.v8.p7). Detailed funding and acknowledgements for the OCAC study are provided in the supplementary material.

Evaluating the role of alcohol consumption in breast and ovarian cancer susceptibility using population based cohort studies and mendelian randomization analyses

Authors

Jue-Sheng Ong*[1,2], Eske M. Derks[1], Mikael Eriksson[3], Jiyuan An[1], Liang-Dar Hwang[4], Douglas F. Easton[5], Paul P. Pharoah[5], Andrew Berchuck[6], Linda E. Kelemen[7], Keitaro Matsuo[8], Georgia Chenevix-Trench[1], Per Hall[3], Stig E. Bojesen[9,10,11], Penelope M Webb[1] and Stuart MacGregor[1].

Affiliations

[1] QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.
[2] Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia.
[3] Medical Epidemiology and Biostatistics, Karolinska Institute, Sweden.
[4] University of Queensland Diamantina Institute, University of Queensland, Translational Research Institute, Brisbane, Queensland, Australia
[5] The Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.
[6] Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina, USA.
[7] Departments of Obstetrics and Gynecology and Public Health Sciences, College of Medicine and Hollings Cancer Center, Medical University of South Carolina, Charleston, South Carolina USA.
[8] Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, Japan.
[9] Faculty of Health and Medical Science, University of Copenhagen, DK-2200 Copenhagen, Denmark
[10] Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, DK-2730 Herlev, Denmark
[11] The Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, DK-2730 Herlev, Denmark.

*Corresponding author

juesheng.ong@qimrberghofer.edu.au

Formatted: Highlight

Formatted: Highlight

Deleted: 6

Abstract

Background: Alcohol consumption is positively correlated with risk for breast cancer in observational studies, but this association may be subject to reverse causation and confounding. The association with epithelial ovarian cancer (EOC) is unclear.

Aim: To evaluate the hypothesis that alcohol consumption is causally associated with the risk to develop of breast cancer and EOC, and subtypes thereof.

Methods: We used observational data from Sweden, Denmark and the United Kingdom to estimate the effect of alcohol consumption on EOC and BRCA risk. We performed a two-sample Mendelian Randomization (MR) analysis combining SNP-cancer summary statistics, with alcohol genetic instruments calibrated using data from UK Biobank. We first evaluated the MR association using a single SNP known to strongly affect alcohol metabolism (rs1229984, *ADH1B*), followed by a multi-SNP approach. We performed stratified MR analysis on estrogen receptor (ER) status for breast cancer and performed additional analyses for high grade serous EOC (HGSOC).

Results: The observational HR for a standard drink/day was 1.06 (95% confidence interval; 1.04,1.08) for breast cancer and 1.00 (0.92,1.08) for EOC. These observational estimates are consistent with previous WCRF findings. ORs using genetically predicted alcohol consumption via rs1229984 genotype were 0.97 (0.88,1.06) for breast cancer and 0.84 (0.69,1.01) for EOC. In subtype analyses, the odds ratio was 0.72 (0.58,0.90) for HGSOC, while stratifying breast cancer by ER status made no meaningful difference for risk estimates. The multi-SNP MR analysis did not reveal a significant influence of alcohol intake on any of the cancer types, including HGSOC [OR 0.91 (0.79,1.05)],

Conclusion: While we reproduced reported observational associations between alcohol intake and increased breast cancer risk, genetic analyses suggest there is little (OR no larger than ~1.06) or no evidence for a causal effect implying that the observational association is confounded by other factors. For overall EOC, observational and MR analyses find no association with alcohol consumption, while for HGSOC, we found evidence for a protective effect of alcohol intake on reduced cancer risk in the single-SNP MR analysis. However, the multi SNP analysis did not reveal a similar protective effect, implicating that ….

Deleted: positively

Deleted: ese are

Commented [ED1]: Why? In the discussion you say that there was no evidence for association in a meta-analysis of 12 studies. Was this study still underpowered?

Deleted: e

Deleted:

Deleted: s

Deleted: Cohort-specific multivariable-adjusted hazard ratios (HR) were computed using age as the underlying time variable. All estimates are for an increase of one standard drink/day.

Deleted: We compared these findings with observational results from those summarised by the World Cancer Research Fund (WCRF).

Deleted: Results are presented as odds ratios (OR).

Deleted: The

Deleted: result was weaker

Deleted: with the multi-SNP approach

Commented [ED2]: Can you speculate on which factors? E.g,. socio economic status?

Commented [ED3]: I think this may be confusing to the reader. Why refer to the upper boundary of the CI? I think your take home message is not clear.

Deleted: studies

Commented [ED4]: I could not remember why you do not report the observational association between alcohol and HGSOC?

Commented [ED5]: But not for the multi-SNP model. You will need to be consistent, do you think there is a causal effect of alcohol on HGSOC or not? You can explain that it is difficult to draw a final conclusion due to the inconsistency between the single and multi-SNP analysis, but I do think you need to make a clear statement.

Deleted: whilst our genetic analysis suggest a potential protective effect on HGSOC.

Introduction

The World Cancer Research Fund (WCRF) concluded that alcohol intake contributes to the risk of breast cancer, with an estimated risk of 9% per 10g/day consumption of ethanol, although there is inadequate evidence to evaluate the association with EOC [1,2]. It is, however, very difficult to measure any effect of elevated alcohol consumption from observational data because of the possible confounding issues: alcohol consumption is itself associated with many other lifestyle and socio-economic factors, which are difficult to quantify, and which might be associated with increased risks of breast cancer and EOC through causal pathways independent of alcohol consumption. Assessing the association in case-control studies is particularly problematic as estimates may additionally be biased by other mechanisms [3], including recall bias due to differences in the accuracy or completeness of the subjective indications of alcohol consumption, selection bias [4,5] against heavy users of alcohol due to the preferential participation of reasonable healthy individuals and reverse causality. The WCRF estimates are based on data from population-based prospective cohort studies, where exposure information is collected before the event of interest occurs, as these are less likely to suffer from bias, but it is still impossible to rule out confounding. Also, such studies typically only measure exposure variables once or a few times, precluding detailed individual modelling of exposures over time [6]. In principle, double blinded randomized trials are the best way to evaluate causality, but such a study would be unethical.

A Mendelian randomization (MR) study, using genetic variants associated with alcohol consumption as an instrument, offers a way to test hypotheses of causality, since the genetic variants are less likely to be associated with other known or unknown confounders, and they are not influenced by (pre-)clinical stages of the diseases [7]. Conceptually, MR relies on the random assortment of genetic variants during meiosis to mimic a "natural" randomized trial [8]. Typically, such a genetic instrument only explains a fraction of the variance of the exposure variable, and therefore MR studies need very large numbers of participants to address questions of causality. For alcohol consumption, previous MR studies have used the rs1229984 variant (as this SNP is associated with high levels of acetaldehyde and facial flushing [9]) as a genetic instrument to evaluate the link between alcohol intake and disease outcomes [10–12]. As larger GWAS have identified more risk loci and GWAS of outcomes of interest have increased in size, power has recently become adequate to support meaningful statistical inference [13].

Using large population-based prospective cohort and Mendelian randomization studies, we tested the hypotheses that elevated alcohol consumption is associated with risks of breast cancer and EOC, and subtypes thereof.

Deleted: is a probably

Deleted: cause

Deleted: but that

Commented [ED6]: Make sure you define all abbreviations (I added this one above).

Deleted: for alcohol consumption

Deleted: answer

Commented [ED7]: I would distinguish the associations in observational studies (Aim 1), and the causal influence of alcohol intake on cancer risk (Aim 2).

Methods

Using data from large scale population based cohorts, we evaluated the observational association between self-reported alcohol consumption and risk of breast and ovarian cancer via Cox-regression analyses. Study-specific hazard ratio (HR) estimates were then combined via a fixed-effect meta-analysis (separately for each cancer). For the genetic causality analyses, we performed a two-sample MR to assess whether genetically predicted alcohol consumption is associated with breast/ovarian cancer susceptibility using publicly available consortia data.

Description of observational cohort studies

Data from Copenhagen General Population Study (CGPS) and Copenhagen City Heart Study (CCHS)

The CGPS[14] and the CCHS[15] are two large prospective general population studies from Denmark. For both of these studies, residents from Copenhagen were invited to complete a baseline questionnaire and undergo a physical examination. The questionnaire includes number of alcoholic drinks consumed daily used to derive standard drinks per week (1 standard drink ~ 12g ethanol). Blood samples were also obtained. In total, 69,420 women participated, 60,205 from the CGPS (enrolled between 2003 to 2015) and the remaining 9,215 from the CCHS (enrolled during four examinations from 1976-78, 1981-83, 1991-94, and 2001-03). A total of 2,312 incident breast cancer and 327 EOC were identified. Women with events prior to examination were excluded from the particular analysis. All participants gave written informed consent, and both CCHS and CGPS were approved by the Danish ethics committees (H-KF01-144/01 and KF100.2039/91). Full details on the observational HR analysis in the CGPS and CCHS are provided in Supplementary Methods.

Data from the Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA).

The KARMA study is a large Swedish breast cancer prospective cohort study comprising 70,877 women who attend regular mammographic screening across four hospitals in Sweden [16]. The aim of the project is to identify risk factors for breast cancer such as mammographic density, genetic and lifestyle factors. Information on tumour characteristics, such as ER status, was identified through registers. Self-reported alcohol intake (in grams) estimated via questionnaires was standardised into number of standard drinks/week using a nominal conversion scale of 10g/standard drink. For our HR analysis, we identified 985 incident breast cancer cases and 59,918 healthy controls with non-missing data on confounders. We did not perform the analysis for EOC due to the limited number of cases (n=57). See Supplementary Methods for a complete description.

Commented [ED8]: Do you correct for multiple testing?

Commented [ED9]: This is not clear to me, when exactly is the examination? Is that the study baseline?

Data from the UK Biobank cohort

The UK Biobank (UKB) cohort consists of 502,000 middle-aged individuals recruited from across the United Kingdom [17]. 487,910 individuals passed initial genetic quality control protocols. We identified 215,830 women genetically similar to those of white-British ancestry through ancestral principal component techniques [18]. The UKB records extensive (n>2,000) phenotypes including anthropometric traits, disease status and lifestyle behaviours. Information about cancer diagnosis among the UKB participants was obtained through data-linkage between self-report, hospital records and cancer registries. Individual cancer types were defined based on ICD-10 definitions, as per previous work [19]. After excluding women with a history of cancer (excluding non-melanoma skin cancer) prior to enrolment, the cohort comprised 141,071 white British women. We further removed participants diagnosed with both breast cancer and EOC. Hence, only 4,068 women were diagnosed with breast cancer and 425 with EOC were retained for the analyses. Cox regression was used to obtain hazard ratios for cancer risk per standard drinks/day increase in alcohol consumption. A complete description of the observational HR analysis for alcohol and cancer in the UKB is provided in Supplementary Methods.

Commented [ED10]: From below, I would guess that the definition of a standard drink is 10g/day, but I would include that info in this paragraph as well.

Meta-analysis of observational findings

We then performed an observational meta-analysis of the association estimates combining the UKB results with those obtained from the CCHS+CGPS and KARMA study for breast cancer and EOC. All association estimates were scaled to reflect a one standard drink per day increase (an increase of ~10g/day of ethanol) to facilitate comparison with our MR findings. Estimates were combined under a fixed-effect inverse variance weighted model using the *rmeta* R library. These results were then compared against the existing WCRF findings on both cancers [1,2].

Mendelian Randomization analyses using genetic results

Deleted: Genetic

In this two-sample MR study, we derived instrumental variables for alcohol consumption from the UK Biobank cohort. We then evaluated whether these alcohol-associated instruments were associated with breast/ovarian cancer risk using GWAS summary statistics obtained from the Breast Cancer Association Consortium (BCAC) and the Ovarian Cancer Association Consortium (OCAC). A flowchart illustrating the complete MR procedure is shown in Figure 1.

(Figure 1 here)

Breast and ovarian cancer risk GWAS

The BCAC GWAS summary statistics [20], derived from a total of 122,977 cases and 105,974 controls of European ancestry, were obtained from a publicly available repository (http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray). Among these, 69,501 of the cases were identified to have ER+ breast cancer, and 21,468 cases were ER- breast cancer cases. Participants in the BCAC were recruited from various case-control and cohort studies around the world. BCAC participants involved in the breast cancer GWAS were genotyped via one of these genotyping platforms: (i) custom Illumina iSelect genotyping arrays, (ii) OncoArray or (iii) the iCOGS array. Genotypes were then imputed against the 1000 Genomes Project Phase III reference panels using IMPUTE2 [21]. A full description of the genetic quality control procedures is given elsewhere [20].

The OCAC GWAS summary statistics [22], derived from a total of 22,406 cases and 40,941 controls of European ancestries, were obtained from a publicly available repository (http://ocac.ccge.medschl.cam.ac.uk/). The genotyping platforms used were broadly similar to those used in the BCAC breast cancer GWAS. Top ancestral principal components were fitted as covariates in both the breast cancer and EOC GWAS model to account for the presence of population substructure. Prior to our main analyses, we excluded SNPs that were poorly imputed (INFO<0.6) or had very low minor allele frequencies (MAF<0.01) for both GWAS datasets. It is important to note that the SNP of largest effect on alcohol consumption (rs1229984) was directly genotyped in both of these studies.

Deriving genetic instruments for alcohol consumption (UKB data)

The complete description of how estimated standard drinks per weeks was derived via self-reported consumption of alcoholic beverages is provided in Supplementary Methods. We performed a GWAS on standard alcoholic drinks per week to calibrate genetic instruments that are predictive of self-reported alcohol consumption among white British women in the UKB. We used the software BOLT-LMM [23], a Bayesian linear mixed model GWAS framework to explicitly model the genetic relatedness within the sample. Genetic sex, age and 10 ancestral principal components were included as covariates. We performed 2 separate GWAS: using i) the estimated alcohol quantity and ii) the estimated alcohol quantity in females only. For each alcohol GWAS result, only SNPs that were genome-wide significant and had MAF>0.01 were retained. SNPs were clumped based on LD (r^2

Deleted: ancestries

Deleted: Clumping on the SNPs

Deleted: were then performed

<0.01) and maximum distance of 1000 kb apart to ensure that selected instruments are strictly independent. We identified 72 instruments (including SNP rs1229984) from the combined-sex drinks/week GWAS (Supplementary Table 3). The combined-sex GWAS was used to robustly identify alcohol associated SNPs but in our main MR analysis we adopted SNP effect sizes estimated among females only. In order to ensure that our analyses were protected against weak instrument bias, we only used 34 out of 72 SNPs that were successfully replicated in the female-only alcohol GWAS (p<1e-5 in females).

Two-sample Mendelian randomization analysis

GWAS summary statistics were used to obtain association estimates for genetic predictors of alcohol on cancer outcomes (breast or ovarian cancer) from the respective consortia (BCAC and OCAC). We extracted the SNP-cancer association estimates and minor allele frequency information for each of the 34 alcohol-associated SNP instruments. In the single instrument rs1229984 MR analysis, we combined the SNP-alcohol and SNP-outcome estimates using a Wald-type ratio estimator [24]. Whilst in the multiple instrument MR analyses (34 SNPs), we fitted an inverse variance weighted (IVW) model to obtain a combined estimate of the causal effect inferred via multiple SNPs [25]. For each test, palindromic SNPs with strands that could not be inferred via allele frequency were excluded. We further scaled our MR estimate to reflect a genetically predicted one drink/day increase in alcohol consumption (by multiplying the predicted change in log(OR) of cancer for 1 standard drink/week by 7).

For the MR analysis on breast cancer, we performed stratified analyses based on estrogen receptor (ER) status, whilst for EOC, we subsequently evaluated the association of alcohol with different histotypes including the most common HGSOC histotype. All statistical analyses (including MR sensitivity analyses) were performed in statistical package R using the *TwoSampleMR* library implemented in the MR-Base platform [26].

Sensitivity analyses

We assessed evidence for a non-linear relationship between alcohol and breast cancer or EOC outcomes by evaluating the dose-response relationship over strata of increasing alcohol intake in the observational analyses. For the KARMA study (where ER status was available for breast cancer cases), we performed stratified analyses to evaluate whether the alcohol-breast cancer association differed by ER status and whether the alcohol-EOC association differed by histotype.

For the genetically derived estimates, to ensure that our findings were not biased by violation of the MR assumptions, we repeated our analyses using the following alternative MR models: MR Egger regression [27], weighted median [28] and maximum likelihood model [29]. Deviations of the MR Egger regression intercept from the null for each tested outcome were used to assess evidence of directional pleiotropy. The multi-SNP MR analyses were then repeated using the MR-PRESSO technique [30] which provides adjusted causal estimates after filtering out heterogeneous SNP-outliers and reported alongside the main (IVW) MR results. Funnel plots and leave-one-out MR plots were also provided to evaluate whether the causal estimates were driven by strong outliers. A detailed description of the MR sensitivity analyses is provided in the Supplementary Methods.

Results

Observational association between alcohol consumption and cancer risk

Breast cancer. Alcohol consumption was associated with increased risk of breast cancer in the CCHS+CGPS cohorts with a HR of 1.09 per standard drink/day (95% C.I. 1.05, 1.13), and the Swedish KARMA study with HR 1.07 (0.97,1.19) while the HR in the UKB dataset was lower (HR 1.04 (1.01, 1.07)). Meta-analysing all these estimates yielded an HR of 1.06 (1.04,1.08) for risk of breast cancer per one standard drink/day (Figure 2 upper panel)

Ovarian cancer. In the UK Biobank, higher alcohol consumption was associated with a reduction in risk for cancers in the ovary with an age-adjusted HR of 0.92 (0.85,0.99). Using the multivariable adjusted model (N=62,774, N=172 cases), the log(HR) was unchanged, albeit with wider confidence intervals (adjusted HR 0.92 (0.83,1.03)) due to missing information on covariates. In CCHS+CGPS, the estimated HR (HR=1.07 (0.96,1.20)) was in the opposite direction, but with 95% CIs that overlapped the estimates from the other two studies. Combining both these estimates yielded an meta-analysed HR of 1.00 (0.92,1.08) for the risk of EOC per one standard drink of 10 g alcohol per day increase in alcohol consumption (Figure 2 lower panel).

(Figure 2 here)

MR analyses – exploring the causal influence of alcohol intake on cancer risk

Validating the genetic instruments for alcohol consumption

Deleted: Conventional o

Deleted: nalyses

Deleted: on

Deleted: Observational association between alcohol consumption with breast and ovarian cancer risk¶

Deleted: ¶

Deleted: Instrumental variable analyses - Genetically predicted alcohol consumption

Rs1229984 was significantly associated with consumption of standard drinks/week (p-value=1e-128) in women from UKB, as expected, and explained 0.23% of the variance. Including additional SNPs associated with alcohol consumption increased the percentage of variance explained to 0.92%. Although the higher proportion of variance explained should translate to narrower confidence intervals on the causal ORs from MR, in practice this only occurred after we removed SNPs which passed a heterogeneity test in MR-PRESSO (Figure 3).

Association of genetically predicted alcohol consumption with breast cancer and EOC

Single SNP approach – the effect of rs1229984. For each copy of the allele (T) of rs1229984, alcohol consumption was associated with a decrease of 2.47 standard drinks/week in women. The MR estimates per 7 units increase in genetically predicted weekly alcohol standard drinks (i.e. one standard drink per day) gave odds ratios of 0.97 (0.88,1.06) for breast cancer and 0.84 (0.69,1.01) for EOC (Figure 3). ER status did not modify the association of rs1229984 and breast cancer (Figure 3). However, rs1229984 reduced the risk of HGSOC (OR of 0.72 (0.58,0.90) per genetically predicted 1 standard drink/day increase) indicating that alcohol intake has a protective effect on HGSOC.

(Figure 3 here)

Multiple SNP approach. For a one unit increase in multiple SNP genetically predicted daily alcohol intake (using 34 variants), the odds ratio of breast cancer was 1.03(0.93,1.14) in standard IVW analysis, with a tighter confidence interval when MR-PRESSO was used to discard heterogeneous SNPs (OR 1.00 [0.93,1.08], figure 3). For EOC, the point estimate was less than one, although with relatively wide confidence intervals (OR 0.89 (0.73,1.08)). Stratification by ER status produced essentially unchanged results for breast cancer. For the association between alcohol and HGSOC, the odds ratio was 0.85 (0.68,1.07), with again higher precision using MR-PRESSO (0.95 [0.85,1.06], Figure 3). Genetically predicted alcohol intake was not associated with most EOC histotypes, although confidence intervals were wide due to limited sample size (Supplementary Figure 1).

The comparison of our genetically derived estimate against our new observational findings and the WCRF results for breast cancer and EOC risk is provided in Figure 4.

(Figure 4 here)

Sensitivity analyses

Deleted: The multiple alcohol associated SNPs (including rs1229984) combined explained

Deleted: of the variance in drinks/week among women in UKB

Deleted: explaining more

Formatted: Font: Italic

Commented [ED11]: Sentence is difficult to follow. Rephrase to "The MR analysis indicated that an increase of one standard drink per day did not significantly influence breast cancer risk (OR=0.97, CIXXX) or EOC risk (OR=XX, CI XXX).

Deleted: a decreased risk was observed between

Deleted: and

Deleted: .

The observational HR association between alcohol and breast cancer and EOC for different levels of alcohol consumption indicated no strong evidence for a non-linear relationship (Supplementary Table 4-5). There was limited evidence that the alcohol-breast cancer association differed by ER status in the KARMA study (Supplementary Table 6). Furthermore, the age-adjusted and fully-adjusted models gave similar estimates suggesting minimal evidence for confounding on the factors that were controlled for (Supplementary Table 7).

(Figure 5 here)

The MR scatter plots for both cancers using the original 34 alcohol SNP instruments are shown in Figure 5. Estimates derived from alternative MR methods (after filtering heterogenous instruments) are presented in Supplementary Table 8 and 9 showing that our findings were robust to weak violation of MR assumption, with the MR-Egger intercepts showing no evidence for directional pleiotropy (Supplementary Table 10). It is worth noting that median and mode-based MR estimates show evidence of a protective association between alcohol and EOC though it can be driven by the strong effect sizes captured via rs1229984. In our pleiotropy assessment, we did not observe evidence for an association between our genetic instruments with potential confounders including age at menarche, oral contraceptive use, smoking quantity, coffee consumption and psychiatric traits, except for BMI in the UKB (See Supplementary Table 11 and 12). However, the magnitude of association between rs1229984 and BMI is small that it is very unlikely to have substantially biased our estimates. Moreover, our MR-PRESSO findings were statistically consistent with the IVW estimates for each trait. The distribution of effect sizes around the null across multiple sensitivity MR analyses provide strong support for an overall null or a very weak positive relationship between alcohol and breast cancer or EOC.

Discussion

In this study, we evaluated the association between alcohol consumption and breast and ovarian cancer using conventional observational prospective designs and MR approaches. Risk estimates for breast cancer from the observational findings were slightly higher than those from MR, but with overlapping confidence intervals. Although the confidence intervals are wider on the MR estimates, the MR design is likely to be robust to some of issues which can hamper interpretation of observational studies, such as confounding. Taken together, for breast cancer it appears likely the true effect of alcohol is null or small (up to OR 1.08 for the multi-SNP instrument or 1.06 for the single-SNP instrument). For EOC, the effect appears null, or possibly protective for the HGSOC subtype.

Deleted: were

Commented [ED12]: This sentence does not provide important info on direction of effect and also does not mention that observational association for breast cancer was significant while the MR analysis did not provide evidence for a causal effect.

Commented [ED13]: The word “true” may be confusing. I assume you mean “causal” effect, or “ the effect that is not sensitive to confounders”.

Commented [ED14]: I would not report the upper boundary of the CI, is that a common thing to do?

Commented [ED15]: It will not be clear whether you refer to observational findings or MR.

Comparison with previous literature

Earlier molecular investigations found that alcohol may be implicated in the development of breast cancer, especially ER+ breast cancer, as it modulates estrogen levels. This negative influence of alcohol is supported by a study investigating the link between alcohol intake and percentage of breast density (PBD), postulating a potential relationship between alcohol intake and breast cancer susceptibility via increased PBD [31]. Similarly, many observational findings have found that a higher level of alcohol consumption is associated with an increased risk of breast cancers [32–35]. This was also supported by a large meta-analysis of 27 cohort studies suggesting that even light drinking (<1 drink/day) is associated with an increased risk of breast cancer in women [36]. In our study, we found suggestive evidence from our observational study meta-analysis that increased alcohol consumption is associated with susceptibility for breast cancers albeit the magnitude of association were slightly lower than those reported by the WCRF [2]. However, the direction of effect measured via IVW MR is consistent although the confidence interval of the MR estimate overlapped the null, suggesting that the true causal nature is consistent with either a small or no effect.

The null association between alcohol and EOC was previously shown in the study by Kelemen et al. [37] pooling together data from 12 case-control studies in OCAC, and in other pooled case-control (PMID:22449732) and cohort (PMID:16495916) studies. In contrast, Cook et al.[38], showed that self-reported wine consumption was associated with a reduction in EOC risk. One possible explanation for such association is that the relationship may have been driven by residual confounding with other exposures correlated with socio-economic factors such as educational attainment [37,38]. In our observational analyses, we did not find strong evidence to support a protective association between alcohol and overall EOC, consistent with the WCRF findings [1]. However, MR analyses on specific EOC subtype reveal suggestive evidence that alcohol might be associated with a reduced risk of HGSOC, and this novel discovery would require future replication efforts to evaluate the influence of alcohol on EOC subtypes separately.

Strength and limitations

Our large sample size combining data from various sources allow us to assess the role of alcohol consumption on breast/ovarian cancer with reasonably good precision. The MR approach provides additional evidence to triangulate evidence for causality. Our additional MR analysis using alcohol consumption instruments calibrated only among European women helps protect against biased inferences due to weak instruments [40]. While these SNPs combined explain only a small amount (~0.92%) of variation in alcohol consumption among women (Supplementary Table 3), due to the large

Commented [ED16]: I found the word "negative" confusing since it may indicate a negative effect direction.

Commented [ED17]: Can you say more? Why would this be the case? Are there confounding variables that lead to the association in the observational study? If so, what variables could that be? Socio economic status?

Formatted: Left, Line spacing: Multiple 1.08 li

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Commented [ED18]: Are these samples larger than Kelemen? If not I would start with the two case-control studies and then report the meta-analysis.

Commented [LK19]: Penny's data were included in the OCAC analysis therefore only OCAC should be cited.

Formatted: Highlight

Deleted: and Webb et al.[39]

sample sizes from both OCAC and BCAC, the confidence intervals on our estimates are reasonably precise (three-fold and two-fold wider than those from BCAC and OCAC, respectively).

This study had some limitations. While genetically derived estimates are least likely to be affected by confounding, the magnitude of association between these genetic instruments and estimated standard drinks rely on the accuracy of self-reported data, which may contain self-report bias. In recent years, investigators have used multi-instrument MR experiments due to availability of genetic data on large cohorts. The multi-instrument approach is expected to reduce the confidence intervals on the causal estimates (relative to the inclusion of just a single SNP), although in practice we only found this to be the case when heterogeneous SNPs were discarded using MR-PRESSO (figure 3). In our situation retaining the rs1229984-only estimate can be informative because rs1229984 by itself is by far the strongest and most extensively studied instrument among the SNP set with well-studied biological insights to justify its association with alcohol consumption. Our biological knowledge on how the other SNPs relate to alcohol consumption is more limited although the MR-Egger intercept on EOC did not show any evidence of directional pleiotropy affecting our MR findings for alcohol on EOC (Supplementary Table 10).

While earlier studies have suggested a link between *ADH* genes (e.g., *ADH1B*) and cancer cell growth [41,42], it is unclear whether these association were mainly driven by a change in alcohol consumption. Our pheWAS findings on rs1229984 reveal no strong evidence that the *ADH1B* instrument is associated with phenotypes linked with breast or ovarian cancer that are unlikely to be mediated through alcohol consumption (Supplementary Table 11). However, we cannot exclude the possibility of rs1229984 being associated with other factors related to carcinogenesis unmeasured confounders. We are unable to assess whether our MR causal inference remain consistent when we conservatively excluded rs1229984 from the main analyses, as it resulted in wide confidence intervals on the estimate (rs1229984 being the instrument that explains the highest amount of genetic variance, Supplementary Table 3).

One of the possible explanations for the discrepancies between the observational and MR estimates in the present study might be attributable to bias in the observational study cohorts. Firstly, selection bias might be present for the Copenhagen cohorts if participants are more healthy than non-participants. Our reliance of self-reported consumption data for the observational analyses is

Commented [ED20]: This was unclear to me. You say larger sample sizes from OCAC and BCAC, Cis reasonably precise, and then compare with BCAC and OCAC? Which is a comparison with itself?

Deleted: remains less certain

Formatted: Font: Italic

Formatted: Font: Italic

Commented [ED21]: Do you refer to this in results? I don't think so and would include in results as well.

Formatted: Font: Italic

Commented [ED22]: Rephrase, difficult to follow.

Commented [ED23]: The second paragraph further lists limitations and this is not entirely clear. Can you number them? Firstly,,,, secondly,, etc

Commented [ED24]: This suggests that there are many discrepancies which is not really the case. The findings are largely consistent. For breast cancer, findings are inconsistent, so I would specifically refer to breast cancer. I couldn't remember why you did not explore HGSOC in the observational study.

Commented [ED25]: Sure, but would this lead to discrepancies between observational and MR.

vulnerable to recall error, and the definition of standard drinks might differ across regions, contributing to higher heterogeneity in our exposures. These limitations, however, cannot explain the significant results since such a selection would be expected to pull estimates towards the null. Genetic estimates are conceivably less affected by these biases, but they can be vulnerable to biases in the presence of horizontal pleiotropy. We performed sensitivity analysis based on filtering out SNPs with heterogenous causal effects to reduce the chances of horizontal pleiotropy biasing the estimate, although in practice this made no meaningful changes to the results. Development of genetic instruments based upon subjective information on a stigmatised risk factor like alcohol, might underestimate the strength of the alcohol genetics association, and thereby lead to an inflated estimate from the instrumental variable analysis [43]. However, this potential limitation cannot explain that the genetic risk estimate for breast was weaker than those estimated from observational analyses. It is also unlikely that these differences arise due to the underlying model itself, given that observational HRs and ORs are similar for low population prevalence outcomes (cancers).

Commented [ED26]: Which results do you refer to?

Commented [ED27]: Is that true? Why? Not if there are systematic relations

Commented [ED28]: What do you mean exactly? Although there may be a stigma surrounding alcohol dependence, I am not sure whether this is also true for consumption (especially not in European ancestry populations). Either way, it needs more explanation.

Commented [ED29]: I do not understand what you mean

The MR estimates for alcohol on breast cancer or EOC remain valid under the assumption that alcohol consumption and log(OR) of these disease outcomes have a linear relationship. This is a strong assumption, given previous speculation about a J-shaped relationship between alcohol and other disease outcomes (e.g. cardiovascular diseases) where abstainers are at higher risk similar to those drinking more than moderate amounts [44,45]. Despite our inability to perform MR-by-stratum (evaluating effect of genetically predicted alcohol consumption on risk of disease at various drinking category) due to insufficient sample size, our observational findings show little evidence that the relationship between alcohol intake and these cancers is non-linear. Given that the rs1229984 variant predicts both drinker status and quantity consumed, modelling the MR association within drinkers-only might potentially induce collider bias [46]. Here we present genetic evidence that a subtle change of alcohol consumption (~1 standard drink/day increase) may be associated with a modest decreased risk of HGSOC, but extrapolating changes to larger quantities (e.g. >2 drinks/day) would potentially violate the linearity assumption [47]. Finally, our study could not evaluate evidence for an interaction effect between alcohol polymorphisms and alcohol consumption on cancer risks.

Commented [ED30]: Why would this be important? What do you mean exactly?

Taken together, the results from our observational and mendelian randomization suggest that for breast cancer the effect of alcohol on risk is null or small (up to OR ~1.06-1.08, depending on SNP instrument used). For EOC, the effect appears null, or possibly protective for the HGSOC subtype.

Commented [ED31]: This is the key message of the paper, and I think this should include a much more clear take-home message. What novel findings do you have compared to other studies. What did we learn? Why is this important?

References

1. World Cancer Research Fund. Systematic Literature Review Continuous Update Project Report: The Associations between Food, Nutrition and Physical Activity and the Risk of Ovarian Cancer [Internet]. 2013 [cited 15 May 2019]. Available: https://www.wcrf.org/sites/default/files/ovarian-cancer-slr.pdf

2. World Cancer Research Fund. Continuous Update Project Expert Report: Diet, nutrition, physical activity and cancer [Internet]. 2018 [cited 15 May 2019]. Available: dietandcancerreport.org

3. Boyko EJ. Observational research--opportunities and limitations. J Diabetes Complications. 2013;27: 642–648.

4. Naimi TS, Stockwell T, Zhao J, Xuan Z, Dangardt F, Saitz R, et al. Selection biases in observational studies affect associations between "moderate" alcohol consumption and mortality. Addiction. 2017;112: 207–214.

5. Gmel G. Beneficial effects of moderate alcohol use-a case for Occam's razor? Addiction. 2017;112: 215–217.

6. Rossow I, Norström T. The use of epidemiology in alcohol research. Addiction. 2013;108: 20–25.

7. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21: 223–242.

8. Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. JAMA. 2009;301: 2386–2388.

9. Macgregor S, Lind PA, Bucholz KK, Hansell NK, Madden PAF, Richter MM, et al. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: an integrated analysis. Hum Mol Genet. 2009;18: 580–593.

10. Skaaby T, Kilpeläinen TO, Taylor AE, Mahendran Y, Wong A, Ahluwalia TS, et al. Association of alcohol consumption with allergic disease and asthma: a multicenter Mendelian Randomization analysis. Addiction. 2018; doi:10.1111/add.14438

11. Pierce BL, Kraft P, Zhang C. Mendelian randomization studies of cancer risk: a literature review. Curr Epidemiol Rep. 2018;5: 184–196.

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

13. Burgess S, Malarstig A. Using Mendelian randomization to assess and develop clinical

Formatted: Dutch (Netherlands)

Formatted: Dutch (Netherlands)

Formatted: Dutch (Netherlands)

Field Code Changed

interventions: limitations and benefits. J Comp Eff Res. 2013;2: 209–212.

14. Bergholdt HKM, Bathum L, Kvetny J, Rasmussen DB, Moldow B, Hoeg T, et al. Study design, participation and characteristics of the Danish General Suburban Population Study. Dan Med J. 2013;60: A4693.

15. Aguib Y, Al Suwaidi J. The Copenhagen City Heart Study (Østerbroundersøgelsen). Glob Cardiol Sci Pract. 2015;2015: 33.

16. Gabrielson M, Eriksson M, Hammarström M, Borgquist S, Leifland K, Czene K, et al. Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). Int J Epidemiol. 2017;46: 1740–1741g.

17. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562: 203–209.

18. MacGregor S, Ong J-S, An J, Han X, Zhou T, Siggs OM, et al. Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. Nat Genet. 2018;50: 1067–1071.

19. Ong J-S, An J, Law MH, Whiteman DC, Neale RE, Gharahkhani P, et al. Height and overall cancer risk and mortality: evidence from a Mendelian randomisation study on 310,000 UK Biobank participants. Br J Cancer. 2018;118: 1262–1267.

20. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551: 92–94.

21. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44: 955–959.

22. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet. 2017;49: 680–691.

23. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47: 284–290.

24. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC- InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015;30: 543–552.

25. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res. 2017;26: 2333–2355.

26. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7. doi:10.7554/eLife.34408

27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44: 512–525.

28. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet

Formatted: Dutch (Netherlands)
Formatted: Dutch (Netherlands)
Formatted: Dutch (Netherlands)
Field Code Changed
Formatted: Dutch (Netherlands)
Formatted: Dutch (Netherlands)
Formatted: Dutch (Netherlands)
Field Code Changed
Formatted: Dutch (Netherlands)
Formatted: Dutch (Netherlands)
Field Code Changed
Formatted: Dutch (Netherlands)

Epidemiol. 2016;40: 304–314.

29. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Stat Med. 2016;35: 1880–1906.

30. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50: 693–698.

31. Ziembicki S, Zhu J, Tse E, Martin LJ, Minkin S, Boyd NF. The Association between Alcohol Consumption and Breast Density: A Systematic Review and Meta-analysis. Cancer Epidemiol Biomarkers Prev. 2017;26: 170–178.

32. Romieu I, Scoccianti C, Chajès V, de Batlle J, Biessy C, Dossus L, et al. Alcohol intake and breast cancer in the European prospective investigation into cancer and nutrition. Int J Cancer. 2015;137: 1921–1930.

33. Chen WY, Rosner B, Hankinson SE, Colditz GA, Willett WC. Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. JAMA. 2011;306: 1884–1890.

34. Li CI, Chlebowski RT, Freiberg M, Johnson KC, Kuller L, Lane D, et al. Alcohol consumption and risk of postmenopausal breast cancer by subtype: the women's health initiative observational study. J Natl Cancer Inst. 2010;102: 1422–1431.

35. Zakhari S, Hoek JB. Epidemiology of Moderate Alcohol Consumption and Breast Cancer: Association or Causation? Cancers . 2018;10. doi:10.3390/cancers10100349

36. Choi Y-J, Myung S-K, Lee J-H. Light Alcohol Drinking and Risk of Cancer: A Meta-Analysis of Cohort Studies. Cancer Res Treat. 2018;50: 474–487.

37. Kelemen LE, Bandera EV, Terry KL, Rossing MA, Brinton LA, Doherty JA, et al. Recent alcohol consumption and risk of incident ovarian carcinoma: a pooled analysis of 5,342 cases and 10,358 controls from the Ovarian Cancer Association Consortium. BMC Cancer. 2013;13: 28.

38. Cook LS, Leung ACY, Swenerton K, Gallagher RP, Magliocco A, Steed H, et al. Adult lifetime alcohol consumption and invasive epithelial ovarian cancer risk in a population-based case-control study. Gynecol Oncol. 2016;140: 277–284.

39. Webb PM, Purdie DM, Bain CJ, Green AC. Alcohol, wine, and risk of epithelial ovarian cancer. Cancer Epidemiol Biomarkers Prev. 2004;13: 592–599.

40. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40: 597–608.

41. Seitz HK, Becker P. Alcohol metabolism and cancer risk. Alcohol Res Health. 2007;30: 38–41, 44–7.

42. Seitz HK, Stickel F. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. Genes Nutr. 2010;5: 121–128.

43. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Smith GD, Munafò MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. Econ Hum Biol. 2014;13: 99–106.

44. Zhao J, Stockwell T, Roemer A, Naimi T, Chikritzhs T. Alcohol Consumption and Mortality From Coronary Heart Disease: An Updated Meta-Analysis of Cohort Studies. J Stud Alcohol Drugs. 2017;78: 375–386.

45. Stockwell T, Zhao J, Chikritzhs T. Whether Low Volume Alcohol Use Is Cardio-Protective Is Important for Public Health Policy So the Available Evidence Deserves Critical Analysis: The Authors Respond. J Stud Alcohol Drugs. 2017;78: 392–393.

46. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol. 2018;47: 226–235.

47. Burgess S, Davies NM, Thompson SG, EPIC-InterAct Consortium. Instrumental variable analysis with a nonlinear exposure-outcome relationship. Epidemiology. 2014;25: 877–885.

Figures and Tables

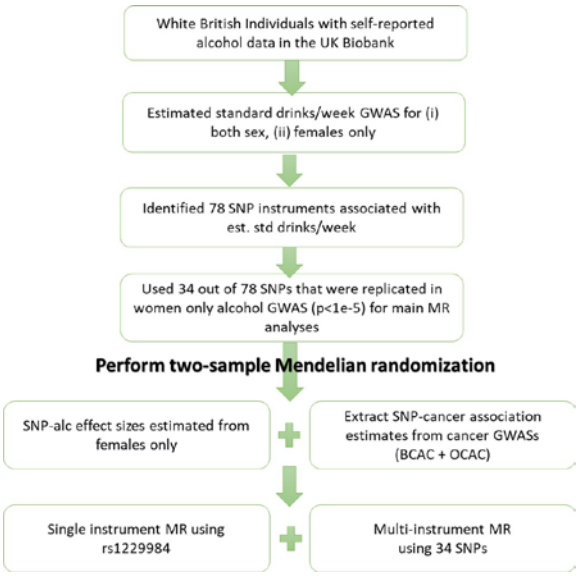
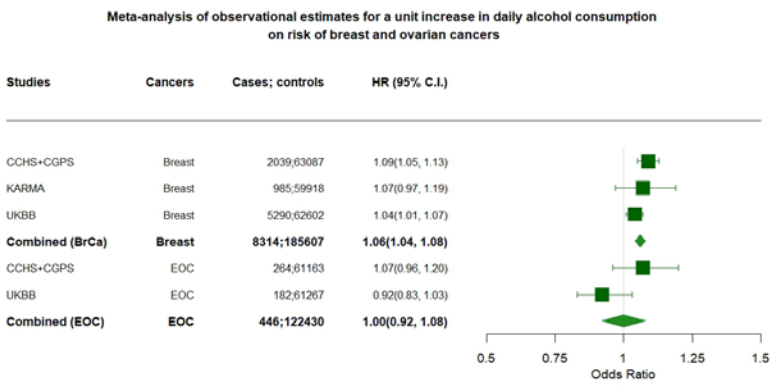


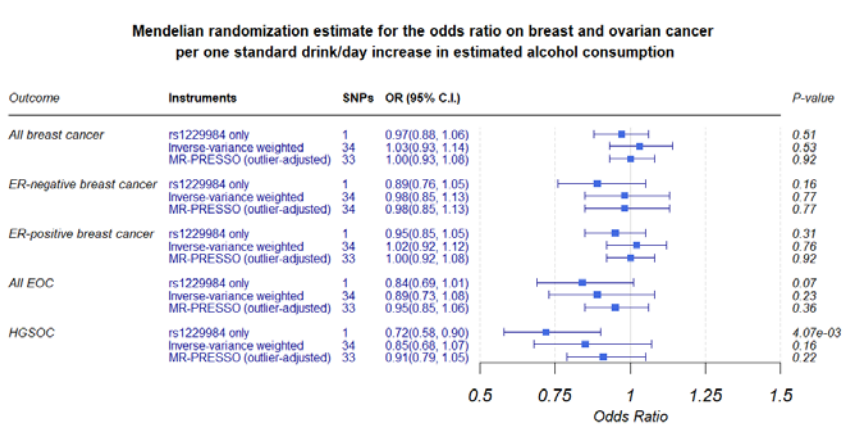
Figure 1. Schematic diagram illustrating the Mendelian randomization (MR) framework for the main analysis.



Commented [ED32]: The vertical line at OR=1 is almost impossible to see, can you increase the thickness of the line? (also in the other similar figures)

Commented [ED33]: Not by subtype as for MR? UKB should have t his information?

Figure 2 Meta-analysis of the observational hazard ratio estimates for daily alcohol consumption on breast and ovarian cancer. Estimates were adjusted for BMI, oral contraceptive use, nulliparity, physical activity and education attainment. Please refer to supplementary table 7 for the estimated HR adjusted for age only.



Commented [ED34]: Do you correct for multiple testing?

Figure 3. Mendelian randomization estimates for the relationship between alcohol consumption and risk of breast/ovarian cancers

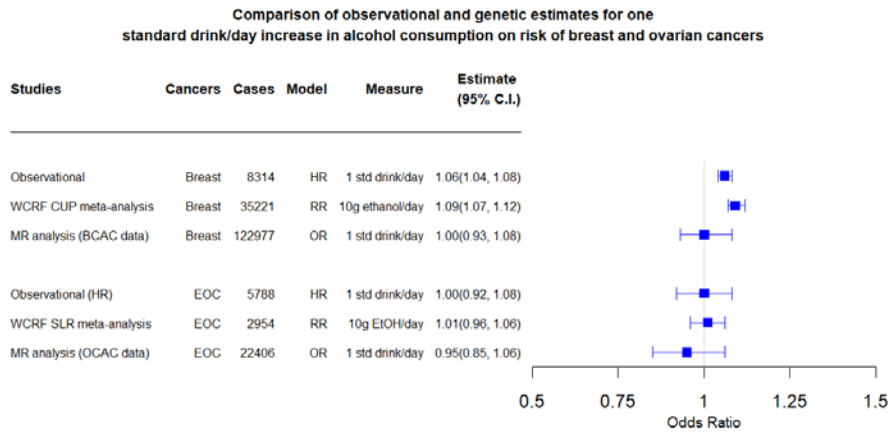


Figure 4. Comparison of observational and genetic (MR) estimates for the association between standard drink per day with breast and ovarian cancer risk. Observational HR estimates were obtained via fixed effect meta-analysis of the studies used in the main analysis. The MR-PRESSO outlier-adjusted estimates were reported here as the MR-analysis findings.

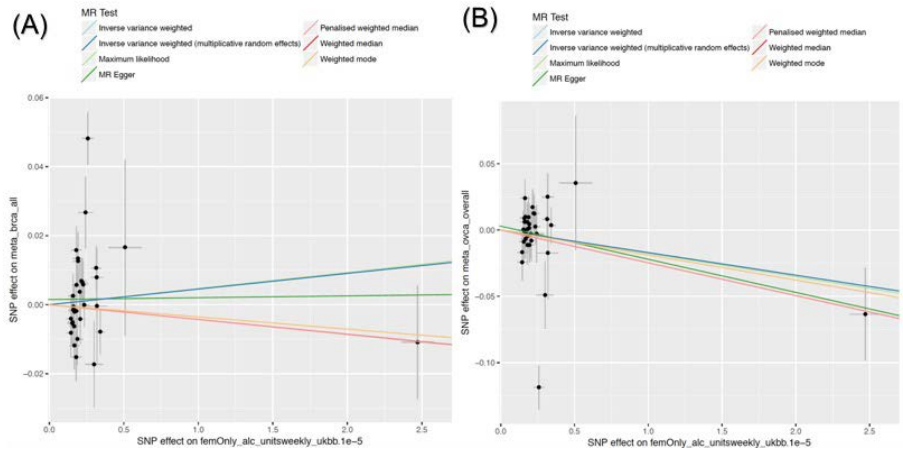


Figure 5. Scatter plot and forest plot for the genetic association between alcohol drinks/week SNP instruments and risk of breast and ovarian cancers. The slope of the fitted line in the scatter plots reflect the MR causal estimates for each type of MR estimator. The forest plot shows the association of a genetically predicted one standard drinks/week increase on log(OR) of the outcome (cancer) risk inferred via each alcohol SNP instrument. The panel (A) refer to the plot for overall EOC; (B) refer to

the plot for the risk of overall breast. For both plots, the left-most point refer to the rs1229984 SNP estimate.