International Journal of Epidemiology, 2019, 1447–1456
doi: 10.1093/ije/dyz144
Advance Access Publication Date: 17 July 2019
Original article

Mendelian Randomization

Association between coffee consumption and overall risk of being diagnosed with or dying from cancer among >300 000 UK Biobank participants in a large-scale Mendelian randomization study

Jue-Sheng Ong [1,2]* Matthew H Law,[1] Jiyuan An,[1] Xikun Han,[1,2] Puya Gharahkhani,[1] David C Whiteman,[3] Rachel E Neale[4] and Stuart MacGregor[1]

[1]Statistical Genetics, Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, [2]Faculty of Medicine, University of Queensland, Brisbane, QLD, Australia, [3]Cancer Control and [4]Cancer Aetiology and Prevention, Department of Population Health, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

*Corresponding author. Statistical Genetics, Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Locked Bag 2000, Herston, QLD 4029, Australia. E-mail: juesheng.ong@qimrberghofer.edu.au

Editorial decision 5 June 2019; Accepted 4 July 2019

Abstract

Background: Previous observational studies have suggested that coffee intake may be associated with a reduction in cancer risk. Mendelian randomization (MR) studies can help clarify whether the observed associations are likely to be causal. Here we evaluated whether coffee intake is associated with: (i) overall risk of being diagnosed with/dying from any cancer; and (ii) risk of individual cancers.

Methods: We identified 46 155 cases (of which 6998 were fatal) and 270 342 controls of White British ancestry from the UK Biobank cohort (UKB), based on ICD10 diagnoses. Individuals with benign tumours were excluded. Coffee intake was self-reported and recorded based on cup/day consumption. We conducted both observational and summary data MR analyses.

Results: There was no observational association between coffee intake and overall cancer risk [odds ratio (OR) per one cup/day increase = 0.99, 95% confidence interval (CI) 0.98, 1.00] or cancer death (OR = 1.01, 0.99, 1.03); the estimated OR from MR is 1.01 (0.94, 1.08) for overall cancer risk and 1.11 (0.95, 1.31) for cancer death. The relationship between coffee intake and individual cancer risks were consistent with a null effect, with most cancers showing little or no associations with coffee. Meta-analysis of our MR findings with publicly available summary data on various cancers do not support a strong causal relationship between coffee and risk of breast, ovarian, lung or prostate cancer, upon correction for multiple testing.

Conclusions: Taken together, coffee intake is not associated with overall risk of being diagnosed with or dying from cancer in UKB. For individual cancers, our findings were not statistically inconsistent with earlier observational studies, although for these we were unable to rule out a small effect on specific types of cancer.

Key words: Mendelian randomization, causal inference, cancer, coffee, instrumental variable, UK Biobank, cancer mortality

Key Messages

- Earlier observational findings suggested a possible association between coffee intake and cancer outcomes.
- Using both observational and Mendelian randomization analysis on more than 300 000 White British participants, we showed that daily coffee intake is not associated with overall cancer risk or cancer death.
- Observational analyses on individual cancers reveal weak evidence that coffee intake has a larger effect on modifying risk of certain cancers; however, larger sample sizes will be required to draw more robust MR inferences on individual cancers.

Introduction

Dietary factors have long been speculated to be associated with cancer risk.[1] These factors include coffee, one of the most widely consumed beverages worldwide. Coffee contains a complex mixture of bioactive ingredients, including substances such as caffeine and kahweol, shown to display anti-tumour effects in animal studies.[2,3] However it is unclear whether or not coffee consumption has anti-cancer effects, as studies to date have produced conflicting findings for overall cancer risk[4–6] and for individual cancers such as breast[7,8] and prostate[9–11] cancer. The differences are partially attributable to variation in study designs (e.g. case-control vs cohort studies), the investigated types of cancers, and types of coffee. Bias arising from unmeasured confounding and reverse causality further complicates causal inferences. Randomized clinical trials (RCTs) are arguably the gold standard to infer causality, although no RCTs have been conducted to date investigating the relationship between coffee and cancer risks.

A complement to traditional epidemiology is to conduct Mendelian randomization (MR) analyses. MR is an instrumental variable-based approach,[12] where the likelihood of causality between an exposure of interest and the outcome, subject to various assumptions, can be inferred using genetic variants associated with the exposure as instruments. As the inheritance of genetic variants is randomized at meiosis, the association between the alleles and the outcome are less likely to be affected by reverse causality and confounding.[12] Coffee consumption is heritable, and MR has been used to show there is no strong evidence for causal links between coffee consumption and ovarian[13] and prostate cancers.[14]

In the present analyses, we leveraged the large sample sizes present in UK Biobank (UKB) to answer the question 'will changing lifetime coffee consumption alter the risk of being diagnosed with or dying from cancer?' by investigating the association between coffee consumption and overall cancer risk and mortality in the UKB cohort. For cancers with a sufficient number of cases ($n > 1000$ as a pragmatic cut-off), we performed secondary analyses to evaluate the association between coffee and individual cancer susceptibility.

Methods

The UK Biobank resource

The UK Biobank cohort study[15] is a population-based cohort study consisting of approximately half a million participants recruited across the UK from 2006 to 2010. This cohort comprises primarily middle-aged participants of European ancestry, ranging from 37 to 73 years old. Participants were required to complete a series of baseline questionnaires, undergo various physical assessments and report medical conditions. To date, the UK Biobank has recorded more than 2000 traits including anthropometric measurements, clinical diagnoses and self-reported behavioural outcomes. We restricted our analyses to 438 870 White British participants with adequate genetic data. Description of genotyping procedures and data cleaning can be found in Supplementary Methods 1, available as Supplementary data at *IJE* online.

In the UKB cohort, 408 191 White British participants had self-reported information on daily coffee consumption.

Coffee consumption was collected as part of the UKB diet survey, where participants were asked 'How many cups of coffee do you drink daily?' (UKB Data-field ID: 1498). Coffee consumption lower than one cup/day were coded as zero. The type of coffee considered includes decaffeinated coffee, instant coffee, ground coffee and any other type of coffee (UKB Data field ID: 1508). For individuals reporting consumption data across multiple visits to the assessment centre, the average was used.

The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee (reference number 06/MRE09/65), and at recruitment all participants gave informed consent to participation in UK Biobank and to being followed up, using a signature capture device.

Observational analyses between coffee intake and cancer outcomes

Using self-reported coffee intake available in the UKB, we evaluated the observational relationship between self-reported cup/day coffee intake (UKB field ID: 1498) and cancer outcomes through logistic regressions. Cases with cancer diagnosis preceding recruitment (prevalent cases) were excluded. We fitted a logistic model adjusting for age, sex, 10 ancestral principal components only and other potential confounders (see Supplementary Table 1, available as Supplementary data at *IJE* online, for complete description). For the analyses of cancer outcomes among women, we used a separate model that additionally accounted for female specific risk factors (Supplementary Table 1). The statistical package R was used for analysis.

Instrumental variable analyses

In brief, we performed an MR analysis to evaluate whether genetic predisposition to coffee consumption was associated with cancer outcomes. We first performed a genome-wide association study (GWAS) of coffee consumption among cancer-free individuals in the UKB cohort, to identify single nucleotide polymorphism (SNP) instruments for coffee consumption. We then evaluate whether these coffee instruments were associated with cancer outcomes. As a robustness check, the MR analyses were also repeated using instruments directly curated from another (independent) coffee GWAS study [the Coffee and Caffeine Genetics Consortium (CCGC) study].

Genetic instrument for coffee consumption

We performed a self-reported coffee intake GWAS (number of cups/day) to identify potential instruments for habitual coffee consumption. First, individuals previously diagnosed with cancer were excluded to avoid bias in MR estimates due to reverse causality, retaining 179 954 males and 212 119 females of White British ancestry for the GWAS analysis. The BOLT-LMM software[16] was used to model the genetic association accounting for cryptic relatedness in the sample (see Supplementary Methods 2, available as Supplementary data at *IJE* online). Before the MR analyses, SNP instruments were clumped [linkage disequilibrium (LD) between instruments <0.01 at a 10-mb window] and selected based on the following criteria: (i) $P<1e\text{-}8$; (ii) minor allele frequency (MAF) >0.05; (iii) SNPs are non-palindromes (A/T and G/C polymorphisms removed if strand cannot be inferred via allele frequencies).

Genetic association of coffee SNPs with cancer diagnosis and cancer death

There were approximately 46 531 individuals diagnosed with cancer recorded in the UKB cohort. We collated information on cancer outcomes in UKB based on various sources: cancer registry and hospital and health records, using the criteria outlined from previous work.[17] Definition of cases, inclusion criteria and handling of sample relatedness are described in Supplementary Methods 3, available as Supplementary data at *IJE* online. Finally, we retained 46 155 cases and 270 342 controls.

Using similar definitions, we obtained 270 342 healthy controls and identified 6998 people who had died from cancer (based on ICD10) for the cancer mortality analysis. The SNP association analyses for cancer risk and mortality were performed as logistic regressions in PLINK2.0 alpha.[18] Covariates included age, sex and the first 10 genetic principal components provided by UK Biobank.

Statistical analyses

Subject to various MR assumptions,[19] a genetic association between the coffee consumption SNPs and cancer outcome would suggest a causal relationship. Variance explained by coffee SNP instruments were calculated (Supplementary Methods 3). We then performed power calculations to evaluate our statistical power to detect effect sizes at various magnitude (Supplementary Table 2, available as Supplementary data at *IJE* online). Moreover, the proportion of coffee non-drinkers in our total sample is 21% (with a very similar distribution for cases and for controls) and hence the presence of non-drinkers is unlikely to affect the overall power of the MR analyses. Here we used the inverse-variance weighted (IVW) Wald-type estimator[20] to derive an MR estimate using the SNP-coffee (cups/day) and the SNP-outcome [log(OR)] associations. The standard error of our IVW estimates were computed using the delta-approximation.[20] Our MR OR estimates reflect the change in cancer risk (or death) for a one cup/day increase

in genetically predicted coffee consumption. MR analyses were also performed for individual cancers with sufficient cases (Supplementary Methods 5, available as Supplementary data at *IJE* online). Publicly available GWAS summary data were available for the risk of ovarian[21] (OCAC), lung[22] (ILCCO), breast[23] (BCAC) and prostate[24] cancers (PRACTICAL), each from their respective consortia. The data for OCAC, BCAC and prostate cancer were obtained directly from their respective online repository, whereas the data for lung cancer from ILCCO were downloaded from the MR-Base[25] platform [www.mrbase.org]. To further improve the precision of our individual cancer risk IVW estimates, these consortia-based estimates were extracted and meta-analysed with our UK Biobank findings on those cancers (Supplementary Methods 6, available as Supplementary data at *IJE* online).

We also assessed potential interaction of coffee intake and smoking on overall cancer outcomes (power for individual cancers was too low), in view of previous work.[26,27] This was done by performing a stratified MR analysis of coffee on cancer outcomes by smoking status (smokers vs never smokers). The MR analyses were performed using the Mendelian randomization R package[28] and the TwoSampleMR R package curated by MR-Base.[25]

Sensitivity analyses

We performed a series of sensitivity analyses to test for potential biases due to ascertainment bias in the UKB,[29] violation of MR assumptions (e.g. SNP pleiotropy) influencing our causal estimates. These include repeating our MR analyses using alternative MR methods, applying a more stringent criterion for instrument selection or alternative SNP instruments from the Coffee and Caffeine Genetics Consortium (CCGC)[30] coffee GWAS (namely rs2470893, rs6968554, rs17685, rs7800944, rs1260326 and rs1481012), assessing heterogeneity in causal estimates and evaluating evidence of horizontal pleiotropy via phenome-wide association studies. These sensitivity analyses are described in greater detail in Supplementary Methods 7, available as Supplementary data at *IJE* online.

Results

Observational findings in UKB

The observational estimates for coffee intake (cup/day) and overall cancer outcomes are shown in Figure 1. The estimated association between cups/day coffee intake and overall cancer risk was very close to null, with relatively tight intervals (OR = 0.99, 95% CI 0.98, 1.00). Our sex-stratified analyses reveal no strong evidence that our estimates differed by sex. There was very weak evidence for an association between coffee intake and cancer mortality (OR = 1.01, 0.99, 1.03).

For individual cancers, the crude observational estimate for coffee and cancer risk adjusting for only ancestry, age, sex and Townsend deprivation (TDI) suggests a potential protective effect for coffee on colorectal, prostate and lung cancers. These associations were removed upon adjusting

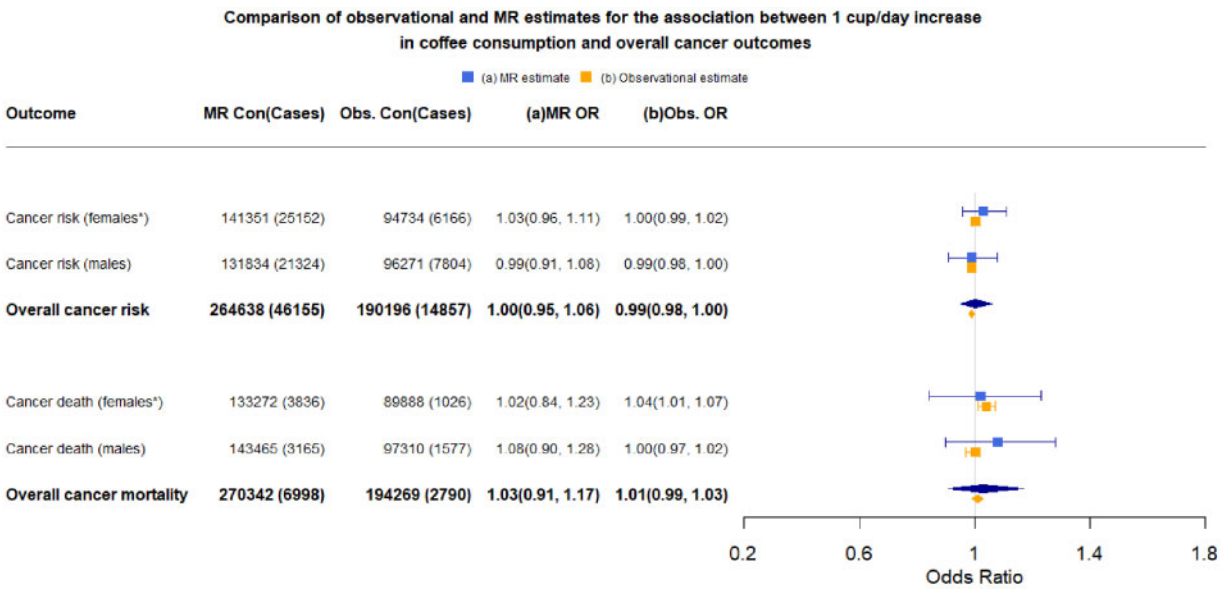

Figure 1. Association between coffee intake with overall cancer risk and cancer mortality in the UK Biobank White British cohort. MR Con(Cases) here refers to the numbers of controls and cancer cases used in the MR analyses. Observational estimates were obtained among incident cases only. For analysis of cancer risk and death in females, the model incorporates additional covariates adjusting for female-specific risk factors (see Supplementary Tables 1 and 3, available as Supplementary data at *IJE* online, for full description of covariates used in observational analyses and estimates under alternative fitted models). The sample size quoted reflects individuals with non-missing data on the covariates after removing related individuals in each study.

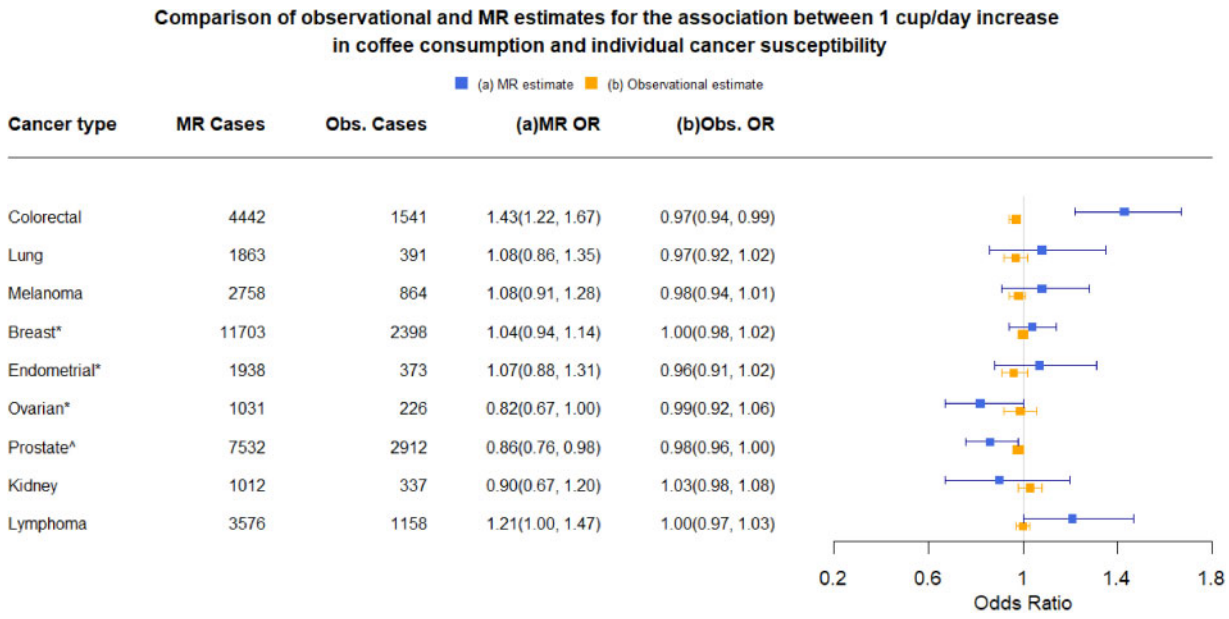

Figure 2. Association between coffee intake and individual cancer susceptibility in the UK Biobank White British cohort. MR Cases here refers to cancer cases used in the MR analyses. MR estimates for all cancers are derived using random effect IVW after filtering out SNPs with strong evidence of heterogeneity. Observational estimates were obtained among incident cases only (Obs. cases refer to cases in the observational analyses). See Supplementary Table 1, available as Supplementary data at *IJE* online, for number of controls used and full description of covariates used in observational analyses. Estimates under alternative fitted models are provided in Supplementary Table 4, available as Supplementary data at *IJE* online. *Estimated in females only, adjusted for female specific risk factors (age of menarche, menopausal status, use of contraceptives). ^Adjusted for smoking quantity (smoking pack-years, tobacco use). Numbers of cases quoted here reflect the numbers after removal of related individuals.

for educational attainment and smoking (pack-years). In the adjusted model, the associations between coffee and cancer risk for most cancers were small (average OR = ~0.98 for one cup/day increase in coffee intake). Results for the individual cancer observational analyses are shown in Figure 2 alongside the MR estimates.

In the UKB cohort, coffee consumption was inversely correlated with tea intake (r = −0.3) but the adjustment for tea intake made no meaningful difference to the results. For potential confounding with smoking, only the association estimate for lung cancer changed upon adjustment for smoking quantity, as the causal estimate attenuates towards the null (OR = 0.97, 0.92, 1.02). Due to the strong evidence of confounding by smoking on our lung cancer estimates, we report the smoking quantity-adjusted estimate between coffee and lung cancer in Figure 2.

Mendelian randomization estimates

For the MR analyses, we identified 35 independent genetic variants robustly associated with coffee intake (cup/day) in the UKB cancer-free samples (Supplementary Tables 5 and 6, available as Supplementary data at *IJE* online). We used the IVW model to obtain a combined estimate of the causal association between coffee and cancer outcome. SNPs that showed strong evidence of heterogeneity in the causal estimates (for each outcome trait separately) were removed and the MR analyses were repeated. Hence, the reported MR estimates in the main analyses reflect the heterogeneity-adjusted estimates (see Figure 1), with the original forest plots provided in Supplementary Figures 1 and 2, available as Supplementary data at *IJE* online. The OR for overall cancer risk for every unit increase in genetically predicted cup/day of coffee intake was 1.01 (95% CI 0.94, 1.08) and the OR on cancer mortality was 1.11 (0.95, 1.31). For both the estimates on cancer risk and those on cancer mortality, we did not observe any strong evidence of a sex difference (Figure 1). Across both cancer risk and mortality, none of the estimated MR associations showed strong evidence for an association at an alpha-threshold of 0.025.

We calculated the MR association between cup/day coffee consumption and susceptibility to nine common cancer types in the UKB cohort (Figure 2). There was weak evidence for an MR association between coffee intake and individual cancer risk in the UKB upon multiple testing corrections (nine cancers), except for colorectal cancer with OR 1.43 (1.21, 1.66). However, the coffee-colorectal cancer association was attenuated towards the null, with OR 1.15 (0.92, 1.46), when only the six coffee SNP instruments (replicated in the CCGC GWAS) or the set of stringent instruments ($P < 1e\text{-}9$) were used (OR 1.11, 0.92, 1.33), revealing limited evidence for a strong association.

Table 1. Meta-analysis of MR association between one cup/day increase in coffee intake and sex-specific cancer risks combining UKB cancer estimates with those from publicly available consortia data

Cancer type	UKB controls (cases)	UKB OR (95% CI)	Group	Group controls (cases)	Group OR (95% CI)	Combined controls (cases)	Meta OR (95% CI)
Breast	141 351 (11 703)	1.04 (0.95, 1.14)	BCAC	105 974 (122 977)	0.93 (0.82, 1.05)	247 325 (134 680)	1.00 (0.92, 1.07)
Ovarian	141 351 (1031)	0.82 (0.67, 1.00)	OCAC	22 406 (40 941)	0.92 (0.79, 1.07)	182 292 (23 437)	0.88 (0.78, 0.99)
Prostate	131 834 (7532)	0.86 (0.75, 0.98)	PRACTICAL	61 112 (79 194)	0.96 (0.84, 1.08)	192 946 (86 726)	0.91 (0.83, 0.99)
Lung	264 638 (1863)	1.08 (0.86, 1.35)	ILCCO	15 861 (11 348)	1.06 (0.89, 1.27)	280 499 (13 211)	1.07 (0.93, 1.23)

GWAS summary statistics for each of the cancer types above were obtained from their respective consortia (Supplementary Methods 6, available as Supplementary data at *IJE* online). Meta OR refers to the fixed effect meta-analysed MR OR.

Table 2. MR association of coffee on overall cancer outcomes stratified by smoking status

Outcome	Among never-smokers		Among ever-smokers		ChiSquare of diff.	*P*-value of difference
	OR	Pval	OR	Pval		
Cancer risk in Females	1.031 (0.910, 1.169)	0.63	1.019 (0.928, 1.120)	0.69	0.02	0.88
Cancer risk in Males	0.978 (0.832, 1.148)	0.78	1.001 (0.898, 1.115)	0.99	0.06	0.81
Cancer risk (Both sexes)	1.006 (0.900, 1.124)	0.92	0.998 (0.922, 1.081)	0.96	0.01	0.91
Cancer mortality in Females	0.926 (0.667, 1.286)	0.65	1.243 (0.941, 1.641)	0.13	1.79	0.18
Cancer mortality in Males	0.985 (0.652, 1.489)	0.94	1.131 (0.887, 1.442)	0.32	0.32	0.57
Cancer mortality (Both sexes)	0.952 (0.733, 1.236)	0.71	1.179 (0.976, 1.425)	0.09	1.69	0.19

Pval here refers to the association estimates at alpha-significance threshold of 0.05. ChiSquare of difference is a chi-square test statistics to evaluate whether the estimates between ever and never-smokers are different; *P*-value of difference is the associated *P*-value.

Meta-analysis of UKB cancer MR estimates with publicly available data from cancer consortia

We performed a meta-analysis of the MR ORs estimated from the UKB samples shown in Figure 2, with ORs derived for breast (BCAC), prostate (PRACTICAL), ovarian (OCAC) and lung cancer (ILCCO), using publicly available data from their respective consortia (Table 1). The estimate on breast cancer risk is very close to null (OR 1.00, 0.92, 1.07), and we observed potential associations between coffee and reduced risk of prostate cancer (OR 0.91, 0.83, 0.99) and ovarian cancer (OR 0.88, 0.78, 0.99) but an increased risk for lung cancer (OR 1.07, 0.93, 1.23) although the 95% CI overlapped one. There were no publicly available data for cancer mortality analyses on these cancers.

Interaction between coffee instruments and smoking behaviour on cancer outcomes

We performed stratified MR analyses to evaluate whether the association between genetically predicted coffee intake (cup/day) and cancer outcomes differed by self-reported smoking status. The estimated ORs for the overall cancer outcomes were similar to those for non-smokers. These are shown in Table 2. Unfortunately, power is limited for performing smoking-stratified MR analyses for individual cancer outcomes.

Sensitivity MR analyses

The heterogeneity statistics of MR estimates pre- and post-adjustment (i.e. SNPs with heterogeneity score $Q > 3.84$ were removed) are shown in Supplementary Table 8, available as Supplementary data at *IJE* online, indicating very weak evidence of global heterogeneity after outlier removal. However, this systematic approach erroneously removed the strongest coffee SNP rs2472297 for the MR analysis on colorectal cancer and lymphoma, potentially biasing MR findings. When the analyses were repeated with the CCGC instruments, the estimated effect sizes approximately halved, showing weak evidence for a causal association (Supplementary Table 12, available as Supplementary data at *IJE* online). Separating our cancer cases into incident and prevalent cases also made no meaningful difference to the overall results. The estimated overall cancer risk OR per cup/day increase in coffee intake for incident cancer cases was 1.02 (1.01, 1.03), with 95% CI overlapping those estimated on prevalent cases (OR 0.96, 0.88, 1.05).

Sensitivity MR analyses using alternative MR models yield similar findings, with very limited evidence of directional pleiotropy (Supplementary Tables 9 and 10, available as Supplementary data at *IJE* online). More importantly, the overall cancer MR results evaluated using alternative sets of coffee instruments were highly consistent (Table 3). For

Table 3. MR association between daily coffee consumption (cup/day) and overall cancer outcomes derived using alternative sets of instruments and SNP-coffee estimates

Traits	Original model		P <1e-9 instruments only		CCGC SNPs (coffee effect size in UKB)		CCGC SNPs (coffee effect size from CCGC)	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Overall cancer risk	1.00 (0.95, 1.06)	0.90	1.00 (0.94, 1.06)	0.95	0.99 (0.93, 1.07)	0.85	1.00 (0.93, 1.08)	0.98
Overall cancer mortality	1.03 (0.91, 1.17)	0.66	1.01 (0.88, 1.16)	0.89	1.03 (0.88, 1.22)	0.68	1.05 (0.89, 1.25)	0.56
Cancer risk (females)	1.03 (0.96, 1.11)	0.44	1.03 (0.95, 1.12)	0.43	1.03 (0.94, 1.13)	0.51	1.04 (0.95, 1.15)	0.39
Cancer mortality (females)	1.02 (0.84, 1.23)	0.87	1.04 (0.85, 1.28)	0.71	1.02 (0.80, 1.29)	0.90	1.00 (0.78, 1.29)	0.98
Cancer risk (males)	0.99 (0.91, 1.08)	0.83	0.97 (0.89, 1.07)	0.59	0.96 (0.83, 1.12)	0.63	0.98 (0.84, 1.15)	0.85
Cancer mortality (males)	1.08 (0.90, 1.28)	0.41	1.08 (0.90, 1.31)	0.41	1.06 (0.83, 1.36)	0.62	1.11 (0.87, 1.41)	0.41

The P < 1e-9 instruments only set refer to the set of coffee SNP instruments with a *P*-value <1e-9 on self-reported coffee consumption estimated from the UK Biobank cancer-free individuals. The CCGC sets refer to the SNP instruments derived from the CCGC coffee GWAS. Estimated OR and 95% CI reflect the OR on cancer outcomes (risk/death) per 1 cup/day increase in genetically predicted coffee consumption.

individual cancer analyses, the MR estimates were widely unchanged using alternative instruments, apart from colorectal cancer where the effect estimate attenuated towards the null. Among our coffee SNP instruments, rs2472297 and rs4410790 explained the most of the phenotypic variation tagged by the genetic instruments. Hence, the presence of any association between these two SNPs and confounders may strongly bias our findings. In our PheWAS look-up, coffee-increasing alleles in rs2472297 and rs4410790 were associated with measures of adiposity such as increased body mass index (BMI), reduced height and increased leg fat mass (Supplementary Figure 5, available as Supplementary data at *IJE* online). However, these associations at most explain 0.05% of the trait variance in these traits (see Supplementary Table 13, available as Supplementary data at *IJE* online), suggesting that in the event the association with cancer is confounded by these traits and the bias is likely to be negligible. Additionally, our previous coffee MR study[13] using similar coffee instruments (i.e. SNPs in the CCGC GWAS) also found very weak evidence that these variants are associated with lipid levels, fasting glucose and other confounders of cancer outcome (through horizontal pleiotropy). When we repeated our analyses by excluding the rs2472297 and rs4410790 variants, the point estimates for the causal effect on cancer outcomes were similar but with wider confidence intervals (Supplementary Tables 11 and 12, available as Supplementary data at *IJE* online).

Discussion

In this study, we provided a comprehensive assessment of the association between coffee consumption and overall risk of being diagnosed with/dying from cancer among White British participants in the UKB cohort. Our MR analyses on cancer death indicate a possible (non-significant) increase in risk of dying from cancer for higher genetically predicted coffee intake although the observational estimates were very close to null. In our observational analyses, we found no strong evidence that coffee was associated with risk of any individual cancer evaluated. Taken altogether, we were able to convincingly show the effect between coffee intake and overall cancer outcomes is likely very close to null.

A recent study[31] by Loftfield and colleagues evaluated whether coffee intake was associated with all-cause mortality and cancer mortality in the UKB. However, there are several differences: (i) our study examined only genetically White British participants, whereas Loftfield *et al.*[31] considered all ethnicities;(ii) our study included a formal MR analysis, whereas they only evaluated interaction with genetic scores of coffee; and (iii) we presented estimates both before and after adjustment for tea intake on our association estimates. Finally, our hypotheses were specific to cancer outcomes (both risk and mortality), whereas Loftfield *et al.*[31] investigated the role of coffee consumption on all-cause mortality. The Loftfield *et al.* study showed that the difference between non-drinkers (reference) and >six cups/day coffee drinkers yield a hazard ratio (HR) of 0.8 on cancer mortality, which linearly translates to an HR of ~0.98 for every cup/day increase in coffee intake, with CIs overlapping our estimates.

Several studies[26,27,32,33] in the past suggested that the observed relationship between coffee intake and disease outcomes might had been modified by smoking behaviour. Furthermore, variants in the caffeine-metabolizing gene *CYP1A2* have also been previously shown to interact with smoking status in determining coffee intake,[34] with some studies suggesting an interaction of *CYP1A2* polymorphisms and smoking intake with cancer outcome.[26,27,35] However, our adjusted observational model and MR analyses stratified by smoking did not find strong evidence for an interaction of coffee and smoking on cancer outcomes.

Earlier observational studies examining the link between coffee intake and individual cancer risk have shown

evidence for a protective effect for coffee consumption against colorectal, prostate, endometrial[36] and postmenopausal breast cancers, and adverse effects for ovarian cancer, with some demonstrating a moderate magnitude of association.[7,8] For instance, in Lafranconi *et al.*[36] the relative risk ratio between coffee intake and endometrial cancer was estimated to be 0.80(0.72, 0.89) per four cups increase in coffee intake. If we extrapolate our exposure to reflect a four-cup change in coffee intake, our MR estimate indicates an opposite direction of effect, albeit with very wide CIs: 1.36 (0.52, 3.52). For cancer mortality, a recent observational study[37] consisting of 10 European cohorts revealed a positive association between coffee and cancer mortality specific to females, with a large effect size observed for ovarian cancer mortality. Our observational estimates for risk were broadly similar to theirs, with wider CIs on our MR estimates, but our findings on ovarian cancer were too underpowered to allow any clear conclusions to be drawn.

To strengthen causal inference, our meta-analysis combining data from the UKB cohort and consortia data suggests a minor protective effect of coffee consumption on ovarian and prostate cancer risk and increased risk for lung cancer, although more data will be required to improve precision of the causal estimates. Results from the MR meta-analysis were broadly consistent with our previous MR work,[13] finding very weak evidence for an association between coffee intake and ovarian cancer. Among previous observational studies for breast, lung, prostate and ovarian cancers, most presented limited evidence[38–41] for an association with coffee intake, except for lung cancer where a previous meta-analysis[42] showed a large adverse effect for heavy coffee drinkers compared with individuals with lower coffee consumption. Our combined MR results were broadly consistent with these observational findings.[38–42]

Our MR study has several notable strengths. The use of genetically predicted coffee intake allows re-assessment of the relationship between coffee and cancer outcomes that are less likely to be confounded. Our sample size is large, homogeneous and broadly genetically representative of the White British population. Our coffee SNP instruments were calibrated from a large GWAS that used the same diet questionnaire among UKB participants, which likely translates to higher statistical power for MR, as less heterogeneity in measurement can potentially lead to better identification of genetic instruments.[43] The instruments combined explained close to 1% of phenotypic variance in coffee consumption, higher than with those used in earlier MR studies.[13,14,44–47] Last, our approach to evaluating overall cancer risk/mortality provides a strong overview on whether changing coffee consumption can help reduce overall cancer burden in the population as a whole. Althoughour all cancer approach is by design heterogeneous, this approach captures whether a change in coffee consumption has a 'net-effect' on modifying cancer risk/mortality in a population, assuming that the distribution of cancers is similar to those sampled in the UK Biobank cohort. In contrast, our cancer-specific estimates aim to highlight whether coffee intake largely influences risk of specific cancer, with the caveat that these estimates are less precise.

Our study does have some limitations. In our primary analyses, the use of instruments calibrated within the outcome sample (UKB) is known to bias causal estimates due to overlapping samples, although we calibrated coffee instruments only among controls to reduce Type I error inflation.[48] We later showed that this is unlikely to be a practical concern in our causal inference, as the results were highly consistent when the CCGC[30] instruments were used instead. The concordance is expected, given a majority of the genetic loci associated with coffee overlap (i.e. the SNPs that are replicated in CCGC explained ~60% of the total variance that is tagged by all of the UKB coffee instruments) (Supplementary Table 7, available as Supplementary data at *IJE* online). Second, although our SNPs were robustly associated with coffee consumption, the defined phenotype itself is heterogeneous, which may potentially limit the generalizability of our findings on specific type or preparation procedure (i.e. brewing) of coffee. We were unable to convincingly discriminate the genetic factors predicting coffee consumption separately from tea consumption, given that a subset of the instruments act via caffeine metabolism. This is consistent with earlier genetic studies[49,50] showing that our instruments are also significantly associated with tea intake in the UKB, albeit less predictive, so these instruments are arguably non-specific to coffee. We argue that this does not confound our inference, because a null MR finding between coffee and cancer here would suggest that both tea and coffee do not influence cancer risk/mortality.

Estimates from our MR analysis on individual cancers were less precise as compared with observational findings. Since the evaluated observational association is extremely small (e.g. OR~0.98) for individual cancers, to achieve similar precision for MR would require ~20 times as many cancer cases, or alternatively identification of more instruments (i.e. variance explained by SNPs~20%), both of which are currently unachievable. Furthermore, we are unable to assess non-linearity between genetic coffee intake and cancer, due to limitations on power. Finally, within UKB there may be biases induced via our inclusion of prevalent cases, although our sensitivity analyses stratifying cancer cases by prevalent and incident cases did not have sufficient power to reliably evaluate any potential

differences. In practice, an analysis using only incident cases might be less susceptible to these biases, given larger case accrual. Future work should use consortia data to improve precision of causal estimates when these data become available.

Taken together, we demonstrate that coffee consumption is unlikely to be strongly associated with overall risk of being diagnosed with or dying from cancer, with confidence intervals on our MR estimates tight enough to rule out all but very small effect sizes. We did not observe a sex difference or any difference when stratified by smoking status. For individual cancers, in most cases the results were similar to those for overall cancer, although due to reduced sample sizes it remains possible there are associations for some specific cancers.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work is supported by a project grant (1123248) from the Australian National Health and Medical Research Council (NHMRC). S.M. is supported by an Australian Research Council Future Fellowship. R.E.N. and D.C.W. are supported by fellowships from the NHMRC (Aust). J.S.O. and X.H. received scholarship support from the University of Queensland and QIMR Berghofer Medical Research Institute.

Acknowledgements

We thank Scott Wood and John Pearson from QIMR Berghofer for IT support. This work was conducted using the UK Biobank Resource (application number 25331).

Conflict of interest: None declared.

References

1. Schoenfeld JD, Ioannidis J. Is everything we eat associated with cancer? A systematic cookbook review. *Am J Clin Nutr* 2013;**97**:127–34.
2. Cavin C, Holzhaeuser D, Scharf G, Constable A, Huber WW, Schilter B. Cafestol and kahweol, two coffee specific diterpenes with anticarcinogenic activity. *Food Chem Toxicol* 2002;**40**:1155–63.
3. Schilter B, Huggett A. Effects of the coffee constituents cafestol and kahweol on the expression of xenobiotic metabolising enzymes in the rat liver. *Toxicol Lett* 1994;**74**:75.
4. Alicandro G, Tavani A, La Vecchia C. Coffee and cancer risk. *Eur J Cancer Prev* 2017;**26**:424–32.
5. Poole R, Kennedy OJ, Roderick P, Fallowfield JA, Hayes PC, Parkes J. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. *BMJ* 2017;**359**:j5024.
6. Wang A, Wang S, Zhu C *et al*. Coffee and cancer risk: a meta-analysis of prospective observational studies. *Sci Rep* 2016;**6**:33711.
7. Bøhn SK, Blomhoff R, Paur I. Coffee and cancer risk, epidemiological evidence, and molecular mechanisms. *Mol Nutr Food Res* 2014;**58**:915–30.
8. Jiang W, Wu Y, Jiang X. Coffee and caffeine intake and breast cancer risk: an updated dose-response meta-analysis of 37 published studies. *Gynecol Oncol* 2013;**129**:620–29.
9. Yu X, Bao Z, Zou J, Dong J. Coffee consumption and risk of cancers: a meta-analysis of cohort studies. *BMC Cancer* 2011;**11**:96.
10. Liu H, Hu G-H, Wang X-C *et al*. Coffee consumption and prostate cancer risk: a meta-analysis of cohort studies. *Nutr Cancer* 2015;**67**:392–400.
11. Xia J, Chen J, Xue J-X, Yang J, Wang Z-J. An up-to-date meta-analysis of coffee consumption and risk of prostate cancer. *Urol J* 2017;**14**:4079–88.
12. Katan MB. Commentary: Mendelian randomization, 18 years on. *Int J Epidemiol* 2004;**33**:10–11.
13. Ong J-S, Hwang L-D, Cuellar-Partida G *et al*. Assessment of moderate coffee consumption and risk of epithelial ovarian cancer: a Mendelian randomization study. *Int J Epidemiol* 2018;**47**:450–59.
14. Taylor AE, Martin RM, Geybels MS *et al*. Investigating the possible causal role of coffee consumption with prostate cancer risk and progression using Mendelian randomization analysis. *Int J Cancer* 2017;**140**:322–28.
15. 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 Med* 2015;**12**:e1001779.
16. Loh P-R, Tucker G, Bulik-Sullivan BK *et al*. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* 2015;**47**:284–90.
17. Ong J-S, An J, Law MH *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**:1262–67.
18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**:7.
19. Burgess S, Thompson SG. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. Boca Raton, FL: CRC Press, 2015.
20. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genet Epidemiol* 2013;**37**:658–65.
21. Phelan CM, Kuchenbaecker KB, Tyrer JP *et al*. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* 2017;**49**:680–91.
22. Wang Y, McKay JD, Rafnar T *et al*. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet* 2014;**46**:736–41.
23. Michailidou K, Lindström S, Dennis J *et al*. Association analysis identifies 65 new breast cancer risk loci. *Nature* 2017;**551**:92–94.
24. Schumacher FR, Al Olama AA, Berndt SI *et al*. Association analyses of more than 140, 000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018;**50**:928–36.
25. Hemani G, Zheng J, Elsworth B *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018, May 30. doi:10.7554/eLife.34408.

26. Cotterchio M, Mirea L, Ozcelik H, Kreiger N. Active cigarette smoking, variants in carcinogen metabolism genes and breast cancer risk among pre- and postmenopausal women in Ontario, Canada. *Breast J* 2014;**20**:468–80.
27. Shahabi A, Corral R, Catsburg C *et al*. Tobacco smoking, polymorphisms in carcinogen metabolism enzyme genes, and risk of localized and advanced prostate cancer: results from the California Collaborative Prostate Cancer Study. *Cancer Med* 2014;**3**:1644–55.
28. Yavorska OO, Burgess S. Mendelian Randomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;**46**:1734–39.
29. 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**:226–35.
30. Cornelis MC, Byrne EM, Esko T *et al*.; Coffee and Caffeine Genetics Consortium. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. *Mol Psychiatry* 2015;**20**:647–56.
31. Loftfield E, Cornelis MC, Caporaso N, Yu K, Sinha R, Freedman N. Association of coffee drinking with mortality by genetic variation in caffeine metabolism: findings from the UK biobank. *JAMA Intern Med* 2018;**178**:1086–97.
32. Guessous I, Dobrinas M, Kutalik Z *et al*. Caffeine intake and CYP1A2 variants associated with high caffeine intake protect non-smokers from hypertension. *Hum Mol Genet* 2012;**21**: 3283–92.
33. Grosso G, Micek A, Godos J *et al*. Coffee consumption and risk of all-cause, cardiovascular, and cancer mortality in smokers and non-smokers: a dose-response meta-analysis. *Eur J Epidemiol* 2016;**31**:1191–205.
34. Dobrinas M, Cornuz J, Pedrido L, Eap CB. Influence of cytochrome P450 oxidoreductase genetic polymorphisms on CYP1A2 activity and inducibility by smoking. *Pharmacogenet Genomics* 2012;**22**:143–51.
35. Pavanello S, Fedeli U, Mastrangelo G *et al*. Role of CYP1A2 polymorphisms on lung cancer risk in a prospective study. *Cancer Genet* 2012;**205**:278–84.
36. Lafranconi A, Micek A, Galvano F *et al*. Coffee decreases the risk of endometrial cancer: a dose-response meta-analysis of prospective cohort studies. *Nutrients* 2017, Nov 9. doi: 10.3390/nu9111223.
37. Gunter MJ, Murphy N, Cross AJ *et al*. Coffee drinking and mortality in 10 European countries: a multinational cohort study. *Ann Intern Med* 2017;**167**:236–47.
38. Arthur R, Kirsh VA, Rohan TE. Associations of coffee, tea and caffeine intake with risk of breast, endometrial and ovarian cancer among Canadian women. *Cancer Epidemiol* 2018;**56**:75–82.
39. Sen A, Papadimitriou N, Lagiou P *et al*. Coffee and tea consumption and risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer* 2019;**144**:240–50.
40. Yaghjyan L, Rich S, Mao L, Mai V, Egan KM. Interactions of coffee consumption and postmenopausal hormone use in relation to breast cancer risk in UK Biobank. *Cancer Causes Control* 2018;**29**:519–25.
41. Shafiei F, Salari-Moghaddam A, Milajerdi A, Larijani B, Esmaillzadeh A. Coffee and caffeine intake and risk of ovarian cancer: a systematic review and meta-analysis. *Int J Gynecol Cancer* 2019;**29**:579–84.
42. Xie Y, Qin J, Nan G, Huang S, Wang Z, Su Y. Coffee consumption and the risk of lung cancer: an updated meta-analysis of epidemiological studies. *Eur J Clin Nutr* 2016;**70**:199–206.
43. Sluis S, van der Verhage M, Posthuma D, Dolan CV. Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS One* 2010;**5**:e13929.
44. Kwok MK. Leung GM, Schooling MC. Habitual coffee consumption and risk of type 2 diabetes, ischemic heart disease, depression and Alzheimer's disease: a Mendelian randomization study. *Sci Rep* 2016;**6**. doi:10.1038/srep36500.
45. Nordestgaard AT, Nordestgaard BG. Coffee intake, cardiovascular disease and all-cause mortality: observational and Mendelian randomization analyses in 95 000-223 000 individuals. *Int J Epidemiol* 2016;**45**:1938–52.
46. Nordestgaard AT, Thomsen M, Nordestgaard BG. Coffee intake and risk of obesity, metabolic syndrome and type 2 diabetes: a Mendelian randomization study. *Int J Epidemiol* 2015;**44**:551–65.
47. Zhou A, Taylor AE, Karhunen V *et al*. Habitual coffee consumption and cognitive function: a Mendelian randomization meta-analysis in up to 415, 530 participants. *Sci Rep*;**8**:7526.
48. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016;**40**:597–608.
49. Taylor AE, Davey Smith G, Munafò MR. Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank. *Addiction* 2018;**113**:148–57.
50. McMahon G, Taylor AE, Davey Smith G, Munafò MR. Phenotype refinement strengthens the association of AHR and CYP1A1 genotype with caffeine consumption. *PLoS One* 2014; **9**:e103448.