Original article

Habitual coffee intake and plasma lipid profile: Evidence from UK Biobank

Ang Zhou a, b, Elina Hyppönen a, b, c, *

a Australian Centre for Precision Health, Unit of Clinical and Health Sciences, University of South Australia, Adelaide, Australia
b South Australian Health and Medical Research Institute, Adelaide, Australia
c Population, Policy and Practice, UCL Institute of Child Health, London, UK

SUMMARY

Background & aims: There is evidence that long-term heavy coffee consumption may adversely affect individuals’ cardiovascular disease (CVD) risk. As hyperlipidemia is a well-established contributor to CVD risk, we investigated the association between habitual coffee intake and plasma lipid profile.

Methods: We used data from up to 362,571 UK Biobank participants to examine phenotypic associations between self-reported coffee intake and plasma lipid profiles, including low-density-lipoproteins cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), total cholesterol (total-C), triglycerides, and apolipoproteins A1 and B (ApoA1 and ApoB). Mendelian randomization (MR) analysis using genetically instrumented coffee intake was used to interrogate the causal nature of coffee–lipid associations.

Results: We observed a positive dose-dependent association between self-reported coffee intake and plasma concentration of LDL-C, ApoB and total-C, with the highest lipid levels seen among participants reported drinking >6 cups/day (P linear trend ≤ 3.24E-55 for all). Consistently, in MR analyses using genetically instrumented coffee intake one cup higher coffee intake was associated with a 0.07 mmol/L (95% CI 0.03 to 0.12), 0.02 g/L (95% CI 0.01 to 0.03), and 0.09 mmol/L (95% CI 0.04 to 0.14) increase in plasma concentration of LDL-C, ApoB, and total-C, respectively.

Conclusions: Our phenotypic and genetic analyses suggest that long-term heavy coffee consumption may lead to unfavourable lipid profile, which could potentially increase individuals’ risk for CVD. These findings may have clinical relevance for people with elevated LDL cholesterol.

© 2021 Elsevier Ltd and European Society for Clinical Nutrition and Metabolism. All rights reserved.

1. Introduction

Coffee is one of the most widely consumed beverages in the world [1], with each day an estimated 3 billion cups of coffee consumed worldwide [2]. Coffee consists of thousands of chemical compounds, with potential for both adverse and beneficial health effects to the cardiovascular system.

The largest meta-analysis to date collated evidence of 36 prospective studies totalling 1,279,804 participants and 36,352 cases of cardiovascular diseases (CVDs), observed a U-shaped association suggesting moderate coffee intake as cardio-protective compared to non-drinkers, and no harm for heavy drinking classified as median 5 cups per day [3]. However, given coffee consumption is often one of the first behaviours to be altered when an individual's health status declines [4], comparisons against non-drinkers may be biased [5]. Indeed, reanalysis of the data by altering the reference group from non-drinkers to light drinkers (a strategy that has been employed to mitigate the potential influence of reverse causality in studies on alcohol intake) [6], diminished the possible beneficial effects of moderate consumption, and tentatively suggested a potential adverse effect for heavy drinking [7]. Concerns were further raised using data from the UK Biobank, where those drinking >6 cups/day were seen to have some elevation in CVD risk compared to people drinking 1–2 cups/day [7].

Hyperlipidemia is a well-established risk factor for CVD risk [8], and pharmacologically lowering circulating low density lipoprotein (LDL) cholesterol is beneficial in primary and secondary CVD prevention [9,10]. Examining the relationship between coffee and lipid profile may provide insights on the health effect of habitual coffee
intake on CVD risk. Coffee beans contain lipid soluble diterpenes, with cafestol being a potent cholesterol elevating compound [11–13]. Cafestol is extracted by hot water, and its level in coffee depends on coffee beans and brewing methods, with the highest concentration found in unfiltered boiled coffee brews and negligible amount in filtered or instant coffee [14]. Small to moderate amount of cafestol is present in commercial coffee available in retail outlets [14]. Although existing randomized controlled trials (RCTs) have provided broadly consistent evidence that coffee consumption, in particular unfiltered coffee is associated with unfavourable changes to lipid profile, these studies all have been of relatively short duration (mean, 45 days), and typically administered only one or two dosages of coffee in the treatment arm [15]. In the current study, we used Mendelian randomization (MR) to investigate evidence for causal effects of long-term habitual coffee consumption on serum lipids. This approach uses genetic variants associated on serum lipids. This approach uses genetic variants associated with the exposure of interest to approximate the exposure, and in the absence of horizontal pleiotropy, where variants influence the outcome through pathways other than that via the exposure, MR has the benefit of reducing bias due to confounding and reverse causation [16].

2. Materials and methods

2.1. Participants

The UK Biobank is a large prospective cohort study with over 500,000 participants aged 37–73 years (99.5% between 40 and 69) recruited from 22 assessment centres across the United Kingdom between March 13, 2006 and Oct 1, 2009 with a goal to improve the prevention, diagnosis and treatment of diseases of middle and old age [17,18]. Participants filled in questionnaires to provide broad information on health and lifestyles at baseline survey, provided blood, urine and saliva samples for biomarker and genetic assays, and took part in clinical assessments. A subsample has also provided some assurance that our MR analyses using either instruments have exceeded the recommended threshold of 10, (r² = 0.43% and F statistic = 231). They were taken from the Coffee and Caffeine Genetics Consortium (CCGC) [20], the largest meta-analysis of genome-wide association studies (GWAS) of habitual coffee consumption, which shares no overlap with the UK Biobank. Given strong evidence for pleiotropy for four of the variants in the PhenOScanner search [21] (Supplemental material and Supplemental Table 1), we also conducted the analysis with a second, restricted set including POR, AHR, CYP1A1, and EFCAB5 (r² = 0.43% and F statistic = 413). The exclusion of pleiotropic variants led to a small drop in the explained variance, but an increment in F statistic owing to the reduction in the number of parameters (i.e. variants) in the regression model. F statistics for both instruments have exceeded the recommended threshold of 10, which provides some assurance that our MR analyses using either instrument were less likely to be affected by weak instrument bias [22]. Regression models for calculating F statistics and variance explained by the coffee variants have been adjusted for age, sex, assessment centre, birth location, SNP array and top 40 genetic principal components.
2.4. Plasma lipid concentration

Plasma lipid concentrations were measured using the Beckman Coulter AU5800 (Beckman Coulter Ltd, UK) by enzymatic protective selection analysis (for LDL-C, mmol/L), enzyme immunoinhibition analysis (for HDL-C, mmol/L), CHO-POD analysis (for Total-C, mmol/L), and immunoturbidimetric analysis (for ApoB and ApoA1, g/L) (http://biobank.ndph.ox.ac.uk/showcase/showcase/docs/serum_biochemistry.pdf). For participants on lipid lowering medication at baseline ($N=61,202, Supplemental Material$), we adjusted for their medication use by diving their lipid concentrations by a lipid-specific correction factor (0.68 for LDL-C, 1.05 for HDL-C, 0.75 for Total-C, 0.87 for triglycerides, 0.72 for ApoB, and 1.06 ApoA1) [23]. Sensitivity analysis excluding participants who took lipid-lowering medications produced similar results (Supplemental Fig. 2).

Further, since participants were not required to fast prior to blood sample taken, we performed a sensitivity analysis to examine if fasting status (last food consumption prior to blood sample taken / < 8 h versus ≥ 8 h) could affect the associations of between coffee intake and lipids concentration.

2.5. Covariates

A wide range of covariates was considered with all measures obtained during the baseline assessment. These covered basic demographics (age, sex, and location); anthropometric measures (BMI, and waist circumference); lifestyle factors, including smoking (non-smokers, ex-smokers, current smokers with no information on the type of tobacco that they smoke, cigar/pipe smokers, cigarette smokers <1–5 cigs/day, 6–10 cigs/day, 11–15 cigs/day, 16–20 cigs/day, 21–25 cigs/day, >25 cigs/day), alcohol intake (never, special occasion only, 1 or 2 times/week, 3–4 times/week, ≥ 5 times/week), and intensity of physical activity (light, moderate, vigorous); general health indicators, including self-reported health status (poor, fair, good, excellent), and long-term illness (no, yes). Socioeconomic status was approximated using Townsend deprivation index reflecting area deprivation [24], and education (None or vocational education, CSE (secondary education), A-levels or higher (further education)).

2.6. Statistical analysis

We investigated the association between habitual coffee intake and lipid profile by examining evidence from phenotypic and genetic association analyses. The phenotypic association of self-reported coffee intake with lipid profile were examined by fitting linear regression models, with participants weighed by $1 \text{ kinship coefficient}$ [25] to account for relatedness. The models were adjusted for covariates listed in the covariate section covering demographic, anthropometric, lifestyle, general health, and socioeconomic aspects of participants, and were also controlled for nuisance variables affecting plasma lipid measurements, including fasting time before blood sample was taken, and sample aliquots for measurement (http://biobank.ndph.ox.ac.uk/showcase/showcase/docs/biomarker_issues.pdf).

Coffee-by-coffee-type interaction analysis was carried out among coffee drinkers, and the evidence of interaction was examined by including coffee-by-coffee-type interaction terms in the model and testing if they could improve the model fit. The interaction terms include coffee (cups/day) x instant coffee and coffee (cups/day) x decaffeinated coffee terms, which were added to the model to allow coffee lipids associations to vary between instant and ground coffee and between decaffeinated and ground coffee, respectively. To examine genetic evidence for coffee-lipid associations, we performed 2-sample MR analyses, with variant-coffee and variant-lipids estimates retrieved from the GWAS for habitual coffee intake [20] and from the UK Biobank, respectively (Supplemental Table 2). Regression models for the SNP-lipid associations in the UK Biobank have been adjusted for age, sex, smoking, fasting time before blood sample taken, sample aliquots, birth location, assessment centre, SNP array, and top 40 genetic principal components. We computed the conventional inverse-variance-weighted (IVW) MR estimate.
We observed a positive dose-dependent linear association between self-reported coffee intake and plasma concentration of LDL-C, with a similar pattern for ApoB and total-C ($P_{\text{linear trend}} \leq 3.24 \times 10^{-55}$ for all, Fig. 2). There was a small difference between non-drinkers and others in the HDL-C, with no clear association pattern for ApoB (Fig. 2). Compared to non-habitual drinkers, a decrease in plasma triglycerides concentrations was observed for those reported drinking 3–4, 5–6, and >6 cups/day (Fig. 2). Sensitivity analysis restricting lipid profile data to the first sample aliquot (N up to 325,291) produced near identical association patterns (Results not shown). Further, there was no statistical evidence that fasting status has affected the strength of coffee–lipid associations ($P_{\text{interaction}} \geq 0.58$ for LDL-C, ApoB and Total-C).

When stratified by coffee types, the coffee-LDL-C association appeared to be slightly stronger among people who reported drinking ground coffee in comparison with those with those who reported drinking instant coffee ($P_{\text{interaction}} = 1.11 \times 10^{-06}$, Fig. 3). This difference was also observed in the coffee associations with ApoB and Total-C, while no notable variations by coffee type were seen for HLD-C, ApoA1, or triglyceride concentrations. In 24-h dietary recall, due to overlaps in participants who reported drinking different coffee types (e.g. participants who reported drinking instant coffee could also report drinking other types of coffee such as Espresso, Cappuccino, or Latte), we were unable to perform a formal interaction test to compare the coffee-LDL-C association pattern between coffee types. Nonetheless, the increment in LDL-C associated with coffee intake appeared to be smaller among participants who reported drinking instant coffee compared to those drinking other types of coffee (Supplemental Fig. 3).

### 3.2. Association of genetically instrumented coffee intake with lipid profile

In IVW analyses using 4 CCGC variants, a cup increment was on average associated with a 0.07 mmol/L (95% CI 0.03 to 0.12), 0.02 g/L (95% CI 0.01 to 0.03), and 0.09 mmol/L (95% CI 0.04 to 0.14) higher plasma concentration of LDL-C, ApoB, and total-C, respectively ($P < 0.002$ for all, Fig. 2). Null associations were observed for HDL-C, ApoA1 and Triglycerides ($P > 0.44$ for all, Fig. 2). W-Median, MR-PRESSO and MR Egger produced similar effect estimates, with MR Egger regression, as expected, returning the least precise estimate (Fig. 2). For all lipid biomarkers, IVW estimates using 8 CCGC variants had wider 95% confidence intervals (Fig. 2), and associations with LDL-C, ApoB, and Total-C, which were apparent with 4 CCGC variants, were only picked up with the W-median and W-mode methods which are robust to outliers, but not with the IVW or MR Egger which are sensitive to outliers (Fig. 2).

For the 8-CCGC-variants instrument, excess heterogeneity between ratio estimates of variants was visually evident (Supplemental Figs. 4A–9A), which was further supported by Cochrane’s Q test ($P < 2.45 \times 10^{-33}$ for all lipid biomarkers, Supplemental Table 3) and MR-PRESSO global test ($P < 3.3 \times 10^{-04}$ for all, Supplemental Table 3). Although there was no statistical evidence that MR Egger intercept had deviated from zero, this was likely due to its low statistical power [32], as indicated by wide confidence intervals (Supplemental Table 3). For the 4-CCGC-variants instrument, ratio estimates of variants were reasonably homogenous for LDL-C, ApoB and total-C (Supplemental Figs. 4B, 5B and 8B, Cochrane’s Q test $\geq 0.003$ and MR-PRESSO global test $\geq 0.03$ for all, Supplemental Table 3); although POR was considered as an outlier for LDL-C and ApoB by MR-PRESSO outlier test, removal of it from the analysis had little impact on the results [IVW 0.064 mmol/L (95% CI, 0.038 to 0.089) for LDL-C and IVW 0.016 g/L (95% CI, 0.006 to 0.026) for ApoB].

### 3.3. Sensitivity analyses examining the robustness of findings using the 4-CCGC-variants instrument

We performed additional sensitivity analyses to evaluate to what extent our primary findings for LDL-C, ApoB and Total-C may have been influenced by smoking and alcohol intake, which are potential confounders for the coffee–lipid associations [33]. Restricting genetic analysis to never smokers or never drinkers or adjusting genetic analysis by alcohol intake or BMI produced similar results with those from our primary analysis (Supplemental Figs. 10, 11 and 12). We further performed leave-one-out analysis excluding the CYP1A1 variant, and found that our primary findings for LDL-C, ApoB and Total-C remained unaffected (Supplemental Fig. 13). To evaluate the directionality of our primary findings, we performed the MR Steiger test, which infers directionality by comparing correlation of SNP-exposure with that of SNP-outcome [34]. Coffee intake was identified as the causal factor for LDL-C, ApoB and Total-C ($P_{\text{steiger}} = 2.04 \times 10^{-65}, 8.66 \times 10^{-69}$ and $2.85 \times 10^{-66}$ respectively).
4. Discussion

Using a large prospective study with comprehensive lipid biomarker information, we examined the association of habitual coffee intake with plasma lipid profile, including LDL-C, HDL-C, total-C, triglycerides, ApoA1, and ApoB. In both phenotypic and genetic analyses, we observed that habitual coffee intake is associated with increases in LDL-C, ApoB and total-C, suggesting that long-term heavy coffee consumption may causally lead to unfavourable lipid profiles. Given the well-established CVD-risk-increasing effect of LDL-C [35], our finding may offer an explanation for the coffee-CVD association previously seen in the UK Biobank [7].
In earlier short-term RCTs [15], coffee consumption has been shown to increase LDL-C concentrations. In the UK Biobank, we observed a positive dose-dependent association between self-reported coffee intake and plasma LDL-C and support from the MR analysis for the association when using pleiotropy-robust methods in the context of 8 CCGC variants and when restricting instruments to the 4 CCGC variants. In line with good practice [36,37], we included all variants known to affect coffee consumption.

Fig. 2. Phenotypic and genetic association of habitual coffee consumption with plasma lipids. LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; Total-C, total cholesterol; ApoAI, Apolipoproteins A1; ApoB, Apolipoproteins B; IVW, inverse-variance-weighted MR; W-Median, weighted median MR; W-Mode, weighted mode MR; MR-presso, MR pleiotropy residual sum and outlier test. All models for phenotypic associations were weighed by the kinship coefficient to account for relatedness and adjusted for basic demographics (age, sex, and location), anthropometric measures (BMI and waist circumference), lifestyle factors (smoking, alcohol intake, habitual tea intake and physical activity), general health indicators (self-reported health, and long-term illness), socioeconomic status (Townsend deprivation index, and education), and factors related to lipid measurement (fasting time before blood sample was taken and sample aliquots for measurement).

In earlier short-term RCTs [15], coffee consumption has been shown to increase LDL-C concentrations. In the UK Biobank, we observed a positive dose-dependent association between self-reported coffee intake and plasma LDL-C and support from the MR analysis for the association when using pleiotropy-robust methods in the context of 8 CCGC variants and when restricting instruments to the 4 CCGC variants. In line with good practice [36,37], we included all variants known to affect coffee consumption.
Fig. 3. Phenotypic association of habitual coffee consumption with plasma lipids by types of coffee. LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; Total-C, total cholesterol; ApoAI, Apolipoproteins A1; ApoB, Apolipoproteins B. All models were weighed by 1—kinship coefficient to account for relatedness and adjusted for basic demographics (age, sex, and location), anthropometric measures (BMI and waist circumference), lifestyle factors (smoking, alcohol intake, habitual tea intake and physical activity), general health indicators (self-reported health, and long-term illness), socioeconomic status (Townsend deprivation index, and education), and factors related to lipid measurement (fasting time before blood sample was taken and sample aliquots for measurement).
consumption and used multiple approaches to account for pleiotropy. While analysis using all 8 coffee variants may in principle have higher statistical power, those included variants at near GCKR, BDNF, MLXIPL and ABCG2, which were each associated at genome-wide significant level with at least 19 other non-coffee-related traits (GCKR with 247 other traits, Supplemental Material). Among these, GCKR, MLXIPL and ABCG2 were directly associated with plasma lipid levels [38,39], and BDNF was associated with anthropometric measures [40] and smoking [41], all potential founders of coffee-lipid associations. In the absence of pleiotropy, individual variants would be expected to provide homogenous estimates for outcome associations, proportional to their influence on the exposure [42]. As a direct indication for bias induced by pleiotropy in the 8 variant analyses, we observed excessive heterogeneity between causal estimates from individual variants, and a clear disparity between the IVW estimate (which assumes no directional pleiotropy) and estimates from pleiotropy-robust methods, such as W-median and W-mode. Restricting analysis to the 4 CCGC coffee variants substantially reduced the heterogeneity between variants, improved precision of IVW estimation and the consensus between estimates from different MR approaches. For these reasons, we believe that causal estimates from 8 coffee variants are likely biased by pleiotropic variants, and analysis using 4 variants provides more reliable estimation of causal effect.

Plasma ApoB and total-C levels are ‘composite’ measures reflecting the collective concentration of multiple lipoprotein particles present in the plasma; both encompass LDL and triglyceriderich particles, such as very-low lipoprotein (VLDL) and intermediate lipoprotein (IDL), with total-C additionally including HDL particles [43]. In the UK Biobank, similar to LDL-C, we saw a dose-dependent positive association for plasma ApoB and total-C levels, which is again consistent with our MR analysis. Given that plasma HDL-C (which contributes to total-C level) and triglycerides (which contributes to both ApoB and total-C) exhibit distinctive association patterns, it can be concluded that these associations are primarily, if not entirely driven by LDL-C. Furthermore, in the UK Biobank, we observed that habitual coffee intake is associated with lower triglyceride concentrations and slightly higher HDL-C. However, given these observational associations were not captured by MR analyses and not supported by existing RCTs [15,44], these may not be of clinical importance.

The observed LDL-increasing effect associated with habitual coffee intake may be attributed to cafestol, a lipid-soluble diterpene present in coffee beans [11,15,45]. Cafestol is a very potent cholesterol-elevating compound [11–13], and is mainly present in unfiltered coffee brews, such as Scandinavian boiled coffees, French press, Turkish/Greek coffee, and espresso (which is often the base for other drinks, such as Latte, Cappuccino, Macchiato, and Caffe Americano), with negligible amount found in filtered or instant coffee [14]. In a meta-analysis of RCTs assessing the effects of coffee intake on serum lipid levels, increases in serum lipids were found to be greater in trials of boiled coffee, with little effect seen in trials of filtered coffee [15]. In the UK Biobank, information on cafetiere (i.e. French press) and filtered coffee was available in 24-h dietary recall. However, unfortunately these two coffee types were grouped together in the questionnaire, making a direct comparison (which would be informative of the involvement of cafestol) impossible. Nonetheless, we observed that the magnitude of the coffee-LDL-C association was weaker among participants who primarily drank instant coffee than those who predominantly drank ground coffee, which could be explained by differences in cafestol intakes. The involvement of cafestol in the long-term effects of habitual coffee consumption warrants further investigation in cohorts with better information on coffee intake. This may have potentially important public health implications for CVD risk reduction at the population level, as if it is indeed cafestol that underlies the coffee-LDL-C association, this would inform the practice of filtering coffee prior to consumption. Indeed, in a recent 20-year follow-up study of over 500,000 Norwegians, compared to individuals who drink 1–4 cups of filtered coffee each day, a 11% and 21% increase in CVD death is seen among men and women who drink the same amount of unfiltered coffee, respectively [46].

This is the largest study to date on the effects of habitual coffee consumption on lipid profiles. To establish causal evidence, we triangulate evidence from phenotypic and genetic approaches, and conduct comprehensive analyses using several complementary approaches to evaluate and mitigate effects by potential pleiotropy as discussed above. However, our study also has some limitations. Despite comprehensive adjustments, the phenotypic analyses are prone to residual confounding from unmeasured founders or imprecisely measured covariates (e.g. smoking). True effects may also be diluted by measurement error, due to using self-reported information and as the lipids were measured from samples taken without overnight fasting. However, given the robust dose-dependent associations observed for LDL-C, ApoB and total-C, it is unlikely that these associations are mere artefacts driven by study design. This is further assured by the fact that these phenotypic associations were consistent with our MR findings using genetically instrumented coffee intake, which is less susceptible to study design issues affecting phenotypic analyses, notably confounding and reverse causation. This said, also the results from MR analyses can be confounded in the presence of population stratification. To minimise related effects, we restricted our analysis to White British ancestry, and included extensive adjustments for measures of subtle population structure, including 40 genetic principal components, birth location and assessment centre. A key challenge with MR studies on coffee intake is pleiotropy, and as discussed above, we conducted comprehensive analyses to mitigate related influences on our findings. Despite consistent evidence across MR methods and sensitivity analyses, it is difficult to fully discount pleiotropic effects. For example, while not captured by our search on the PhenoScanner [21], the CYP1A1 gene encodes proteins belonging to the cytochrome P450 superfamily of enzymes, which catalyse reactions involved in drug metabolism and lipid synthesis [47]. However, we conducted leave-one-out analyses excluding all variants one by one and our findings were not affected by the exclusion of this (or other) variant. Our findings support those from shorter term RCTs [15], while long-term trials enabling the investigation of habitual patterns of consumption may be difficult to conduct. Finally, there is some evidence for a healthy volunteer effect in the UK Biobank [48], with related selection of participants potentially leading to collider bias. In principle this could affect both phenotypic and genetic analyses, although empirical evidence comparing risk factor-disease estimations in the UK Biobank with those from 18 nationally representative studies has demonstrated that selection appears to have little impact on exposure-disease associations from UK Biobank [49]. Further, a simulation study with realistic scenarios has also suggested that impact of collider bias caused by selection on MR studies is likely to be relatively modest [50].

In conclusion, in a large prospective cohort, both phenotypic and genetic analysis suggest that higher habitual coffee intake can contribute towards an adverse lipid profile. While the observed effects of coffee were relatively modest, these findings may have clinical relevance for people with elevated LDL cholesterol.

Sources of funding

This study was financially supported by the National Health and Medical Research Council, Australia (GNT1123603). The funder had...
no role in the design, implementation, analysis, and interpretation of the data.

**Authors’ contributions**

EH: conceived the study and designed the research question; AZ analyzed the data and wrote the first draft. Both authors interpreted the results and drafted the manuscript; revised the manuscript critically for important intellectual content and read and approved the final manuscript.

**Conflict of interest**

Authors have no conflicts of interest to declare. National Health and Medical Research Council had no role in the design, implementation, analysis, and interpretation of the data.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clnu.2020.12.042.

**References**


