Investigating causal relationships between coffee consumption and gynecological diseases: a Mendelian randomization study

Man Luo[1], Minghui Li[1,2], Mengsi Zhou[1], Lin Wang[1], Xiaofeng Yang[1]*

[1]Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou, China
[2]Xinxiang Medical University, Xinxiang, China

Submitted: 12 October 2024; **Accepted:** 4 February 2025
Online publication: 5 March 2025

Arch Med Sci
DOI: https://doi.org/10.5114/aoms/200788

***Corresponding author:**
Xiaofeng Yang
Zhengzhou Central Hospital Affiliated to Zhengzhou University
Zhengzhou, China
E-mail:
yangxiaofeng1932@126.com

Abstract

Introduction: Gynecological diseases, including infections, endocrine disorders, and tumors, significantly impact women's quality of life. Emerging evidence suggests that metabolic factors, nutrition, and dietary habits, such as coffee consumption, may influence these conditions. This study employed two-sample Mendelian randomization (2SMR) to investigate the relationship between coffee intake and gynecological diseases.

Material and methods: Publicly available genome-wide association study (GWAS) data from the Integrative Epidemiology Unit (IEU) GWAS database were analyzed using the TwoSampleMR R package. Data on coffee consumption were extracted from the UK Biobank. Instrumental variables were selected based on $p < 5 \times 10^{-8}$ and F-statistic > 10, and causal relationships were assessed using inverse variance weighting (IVW) and the Wald ratio (WR) method. Heterogeneity and pleiotropy were tested using MR-Egger regression. A negative control analysis using skin color was performed to address confounding by population stratification.

Results: The 2SMR analysis identified significant associations between coffee intake and reduced risks of ovarian cancer (OR = 0.07, 95% CI: 0.01–0.84), ovarian cyst (OR = 0.68, 95% CI: 0.49–0.95), and endometriosis (OR = 0.99, 95% CI: 0.99–1.00). Conversely, increased risks were noted for endometrioid ovarian cancer (OR = 7.88, 95% CI: 1.05–59.22). Analysis of different coffee types revealed that ground coffee was associated with decreased risks of ovarian cancer and endometriosis, while decaffeinated coffee showed positive associations with ovarian cancer and endometriosis.

Conclusions: Coffee intake, especially ground coffee, may lower the risk of certain gynecological diseases. However, decaffeinated coffee may increase these risks. Further research is needed to understand the mechanisms behind these findings.

Key words: gynecological diseases, coffee consumption, Mendelian randomization, ovarian cancer, endometriosis, nutritional epidemiology.

Introduction

Gynecological diseases, including infections, endocrine disorders, and tumors, significantly impact women's health and quality of life. Effective treatments for many gynecological diseases are still lacking, and current medications often provide only temporary relief, accompanied by unwanted side effects, while disease may recur when medication is dis-

continued. Understanding the underlying mechanisms and risk factors associated with the onset and progression of gynecological disorders is crucial to minimizing their impact on women's health. Notably, metabolic disorders have been linked to many gynecological diseases, in addition to infections [1]. Extensive epidemiological studies have highlighted the significant influence of nutritional factors and dietary habits in the development of gynecological diseases [2, 3]. Understanding diet's impact on gynecology is key to improving diagnostics and treatments for gynecological diseases.

Coffee, known for its diverse bioactive compounds, has significant effects on long-term human health [4]. Numerous studies have focused on the health outcomes associated with coffee intake, revealing links to various benefits, including reduced overall mortality rates [5] and decreased risks of cardiovascular death [5, 6], liver disease [7, 8], type 2 diabetes [9], and Parkinson's disease [4, 10]. Intervention trials show that coffee, primarily caffeine, enhances short-term glucose metabolism and insulin sensitivity [11, 12]. A meta-analysis showed that each additional daily cup of coffee reduces type 2 diabetes risk by 7% [13]. Caffeine, the key component in coffee, has been reported to influence estradiol levels and is associated with metabolic regulation [14–16].

Different coffee types and preparation methods significantly affect bioactive compounds such as caffeine, chlorogenic acids, and diterpenes. Ground coffee retains more diterpenes than filtered coffee, while decaffeinated coffee has negligible caffeine. Such variations may influence health effects, highlighting the need to study different coffee types and their links to gynecological diseases, as preferences shape intake patterns [17]. Individual preferences for specific coffee types may further contribute to distinct health effects, underscoring the importance of exploring these associations in greater detail.

Studies on coffee's impact on gynecological diseases show conflicting results, with some suggesting protective effects and others reporting no association or harm. These discrepancies often stem from confounding factors, limited epidemiological data, and neglect of variables such as coffee type, preparation method, and caffeine metabolism. Research gaps, particularly regarding long-term effects and dose-response relationships, highlight the need for more robust studies.

Correlation analysis cannot establish causation, as it cannot confirm whether one variable directly causes another or if the relationship is confounded. This is a key challenge in studying coffee's effects on gynecological diseases, where observational studies face issues such as confounding, reverse causality, and self-report biases. While randomized controlled trials (RCTs) are the gold standard for causal inference, their feasibility is limited by resource and ethical constraints. Mendelian randomization (MR) offers an alternative by using genetic variants to assess causality more reliably.

MR offers several significant advantages over traditional observational studies, particularly when investigating causal relationships between exposure factors, such as coffee consumption, and disease outcomes. MR offers genetic variations as instrumental variables, enabling researchers to infer causal associations without the confounding biases typical of observational studies. The core concept of Mendelian randomization was based on the law of independent assortment (Mendel's second law), in which genetic variations are randomly assigned, similar to the random grouping in an RCT [18]. MR avoids behavioral, social, or psychological confounding by using genetic variants, which precede disease outcomes. Furthermore, because genetic variations precede disease outcomes, MR helps establish the direction of causality, providing more reliable insights into how exposures lead to specific health outcomes, rather than the reverse. In this way, MR offers a more robust approach to addressing causal questions that traditional methods cannot definitively answer.

Therefore, in this study, we aimed to address the research gaps regarding the role of coffee consumption in gynecological diseases. Specifically, we conducted 2SMR analysis to investigate the association between overall coffee consumption and the risk of several common gynecological diseases. Additionally, we explored whether different coffee types, including ground coffee, instant coffee, filtered coffee, and decaffeinated coffee, exert differential effects on these diseases. This study seeks to provide novel insights into the causal relationships between coffee intake and gynecological health, contributing to a deeper understanding of how dietary factors influence disease risk.

Material and methods

Ethics statement

This study entailed the analysis of publicly available summary genome-wide association studies (GWAS) data derived from the Integrative Epidemiology Unit (IEU) GWAS database. No direct engagement with human participants or the gathering of fresh data occurred. Consequently, this investigation qualifies as secondary data analysis, thus exempting it from the necessity of obtaining informed consent.

Data preprocessing for exposure and outcome

The IEU GWAS database, which includes 42,346 GWAS summary datasets, was utilized

for conducting 2SMR analysis with the TwoSampleMR R package (version 0.5.6) (https://mrcieu.github.io/TwoSampleMR) [19]. Coffee consumption data were sourced from two surveys within the UK Biobank, specifically Field IDs 1498, 1508, and 100240. Field ID 1498 recorded the number of cups of coffee consumed per day, with participants responding to the question: "How many cups of coffee do you drink each DAY? (Include decaffeinated coffee)". The average daily coffee consumption among participants was 2.13 cups. Additionally, Field ID 1508 captured the types of coffee consumed, including decaffeinated, instant, ground, and other types, with data gathered from participants who reported consuming at least one cup of coffee per day, or less, as indicated in Field 1498. For coffee type data (Field ID 1508), responses from participants who reported "Do not know" or "Prefer not to answer" were excluded. Further, Field ID 100240 recorded whether participants consumed any coffee the previous day, with the question: "Did you drink any coffee yesterday?" This survey included all forms of coffee, including hot, cold/iced, instant, shop-bought (such as Starbucks), machine-made, or hand-made, but excluded substitutes such as Barley Cup or Dandelion Coffee.

The dataset on coffee consumption (UK Biobank, ukb-b-5237), comprising 604,070 records from 501,487 participants, focused on daily coffee intake and excluded responses outside the 0-99 cup range. Responses indicating more than 10 cups per day were verified. Furthermore, entries marked as "Do not know", "Prefer not to answer", or "Less than one" were excluded from the analysis. Regarding coffee type data (Field ID ukb-d-1508), 471,172 records from 393,733 participants were analyzed, with exclusions applied to responses marked as "Do not know" or "Prefer not to answer". Notably, the data on coffee consumption did not include participant age or the duration of coffee exposure, which may limit the scope of interpretation. The sample sizes for all utilized coffee data are presented in Supplementary Table SI, along with a detailed overview of the UK Biobank field IDs and the coffee-related survey questions.

Coffee consumption was treated as a continuous variable, measured as the number of cups consumed per day, and recorded as numeric values. Coffee type was treated as a categorical variable, comprising the following categories: ground coffee, instant coffee, decaffeinated coffee, filtered coffee, and filtered coffee with milk. To reduce confounding, comparisons were primarily conducted between coffee types differing in caffeine content but prepared using similar techniques (e.g., filtered coffee vs. decaffeinated coffee). Variables with multiple simultaneous differences (e.g., filtered coffee with milk vs. decaffeinated coffee) were avoided in the primary analysis to ensure meaningful comparisons. Gynecological diseases, including ovarian cancer, ovarian cysts, endometriosis, and endometrioid ovarian cancer, were treated as binary outcomes, with each condition coded as 1 for presence and 0 for absence based on GWAS-derived summary statistics.

Instrumental variables (IVs) pertinent to coffee were selected based on a p-value threshold of $< 5 \times 10^{-8}$ and an F-statistic > 10, to affirm the robustness and independence of the IVs. Linkage disequilibrium (LD) analysis, employing the European (EUR) genotype from the 1000 Genomes Project as a reference, was conducted to ensure IV independence, setting the maximum LD R^2 value at 0.01 with a search distance of 1000 kb. The methodology of the 2SMR analysis is delineated in Figure 1.

Data harmonization and causal effect evaluation

Data harmonization was achieved using the harmonise_data() function in the TwoSampleMR package [19], ensuring alignment of SNP effects across both exposure and outcome datasets to the same allele. Inverse variance weighting (IVW) was applied to assess causal relationships for multiple genetic variants associated with coffee consumption and types. The Wald ratio (WR) method

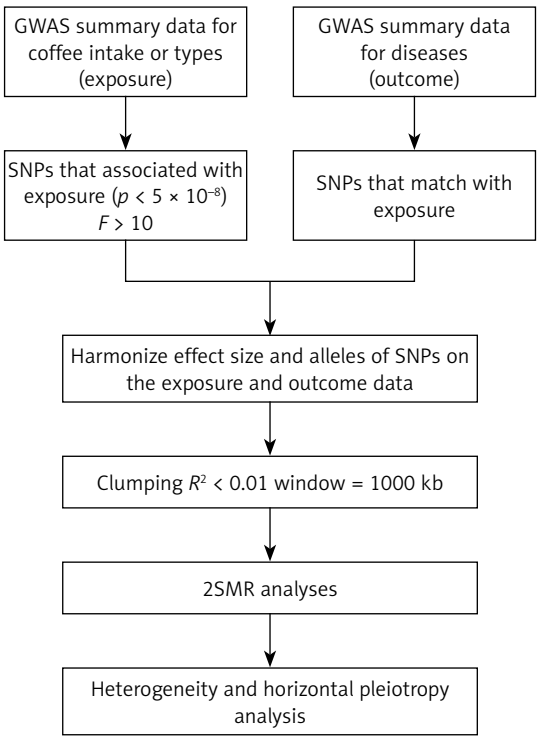

Figure 1. Workflow of two-sample MR analysis

was employed for individual SNP causal effect measurement.

Heterogeneity and horizontal pleiotropy analysis

Heterogeneity in the instrumental variables was assessed using the MR-Egger method, considering a *p*-value > 0.05 indicative of negligible heterogeneity. MR-Egger regression analysis was applied to exposure data (coffee intake or types) involving more than 3 SNPs. The regression intercept from MR-Egger analysis served to evaluate horizontal pleiotropy, with a Q_*p*-value > 0.05 suggesting minimal likelihood of horizontal pleiotropy, thereby allowing its effect to be disregarded.

Population stratification and negative control analysis

To address potential confounding from population stratification in our Mendelian randomization (MR) analyses, a negative control outcome analysis was conducted. Following Liu *et al*. [20], skin color was chosen as the negative control due to its biological independence from coffee intake but potential susceptibility to population structure. The analysis utilized the same MR framework as the primary analysis, ensuring consistent instrumental variable (IV) selection and statistical evaluation. IVs were harmonized using the harmonise_data() function in the TwoSampleMR R package, and causal estimates were calculated using inverse variance weighting (IVW) for multiple SNPs and the Wald ratio for single SNPs.

Data analysis

All analyses were conducted utilizing the TwoSampleMR package (version 0.5.6) in R (version 3.6 or higher), ensuring rigorous evaluation of the data and findings presented.

Results

Exposure and outcome data analysis

The GWAS summary data encompassing 14 traits related to coffee intake and types were identified as exposure variables from the IEU GWAS database (Supplementary Table SI). Conversely, 22 traits pertaining to gynecological diseases (including endometriosis, ovarian cyst, and ovarian cancer) were collated as outcome variables for 2SMR analysis (Supplementary Table SII). The comprehensive workflow of the 2SMR analysis is illustrated in Figure 1.

Causal effects of coffee intake on gynecological diseases

In exploring the nexus between coffee consumption and the incidence of gynecological diseases, 2SMR analysis revealed nominally significant causal relationships between coffee intake and a subset of gynecological conditions including ovarian cancer, ovarian cyst, endometrioid ovarian cancer, and endometriosis (Supplementary Table SIII and Figure 2). MR-Egger regression indicated heterogeneity between coffee intake and the outcomes of ovarian cancer and endometriosis, whereas no significant pleiotropy was detected, allowing for the omission of its impact on causal effect estimations ($p > 0.05$, Supplementary Tables SIV, SV). Forest plots elucidated the odds ratios (ORs) and 95% confidence intervals (CIs) across varying levels of coffee consumption. Specifically, elevated coffee intake was correlated with a diminished risk of ovarian cancer (OR = 0.07, 95% CI: 0.01–0.84), endometriosis (OR = 0.99, 95% CI: 0.9–1.00), and ovarian cysts (OR = 0.68, 95% CI: 0.49–0.95). The association for endometriosis (OR = 0.99, 95% CI: 0.99–1.00) suggests a very minimal effect, with the confidence interval close to

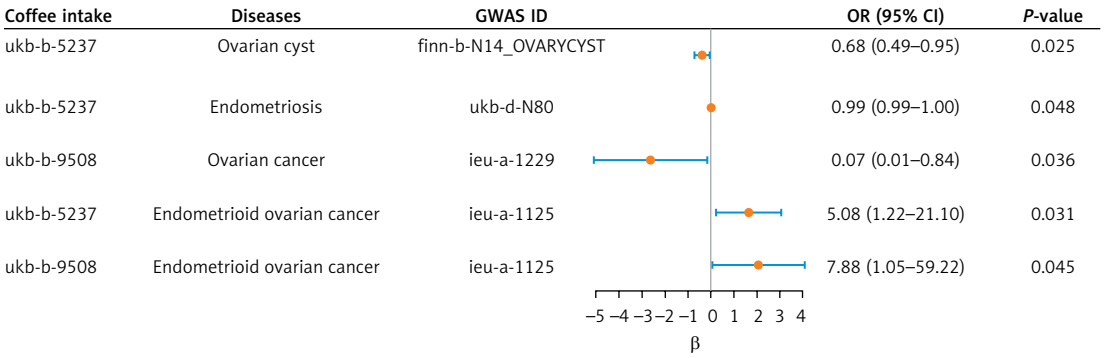

Figure 2. The forest plot illustrates significant findings from the two-sample Mendelian randomization (2SMR) analysis examining the associations between coffee intake and gynecological diseases. Coffee intake was evaluated using two distinct variables: ukb-b-9508, a binary measure representing coffee consumption (consumed vs not consumed), and ukb-b-5237, an ordinal categorical variable capturing coffee intake levels, with odds ratios (ORs) comparing the highest intake category (e.g., ≥ 5 cups/day) to the lowest category (non-drinkers). The figure reports ORs alongside 95% confidence intervals (CIs), with statistical significance defined as $p < 0.05$

1.00, indicating no strong evidence of a significant association. In contrast, coffee intake was associated with a higher risk of endometrioid ovarian cancer, with wide confidence intervals (OR = 7.88, 95% CI: 1.05–59.22; ukb-b-5237: OR = 5.08, 95% CI: 1.22–21.10), suggesting substantial uncertainty in this estimate.

Causal effects of coffee types on gynecological diseases

Further analysis explored the causal relationships between specific coffee types (ground, filtered, filtered with milk, and decaffeinated) and four gynecological conditions. The associations are summarized in Figure 3 and Supplementary Table SVI–SVIII. Ground coffee consumption, treated as a binary variable (consumed or not consumed), was associated with a decreased risk of ovarian cancer (OR = 0.09, 95% CI: 0.01–0.84) and endometriosis of the uterus (OR = 0.98, 95% CI: 0.97–0.99). Conversely, decaffeinated coffee consumption was positively associated with an increased risk of ovarian cancer (OR = 1.05, 95% CI: 1.00–1.10) and endometriosis (OR = 1.02, 95% CI: 1.00–1.03). Filtered coffee and filtered coffee with milk were also associated with reduced risks of endometriosis of the uterus (filtered coffee: OR = 0.99, 95% CI: 0.99–1.00; filtered coffee with milk: OR = 0.98, 95% CI: 0.97–0.99). However, decaffeinated coffee and instant coffee showed potential detrimental effects on endometriosis of the uterus and endometrioid ovarian cancer. These findings suggest that the effects of coffee on gy-

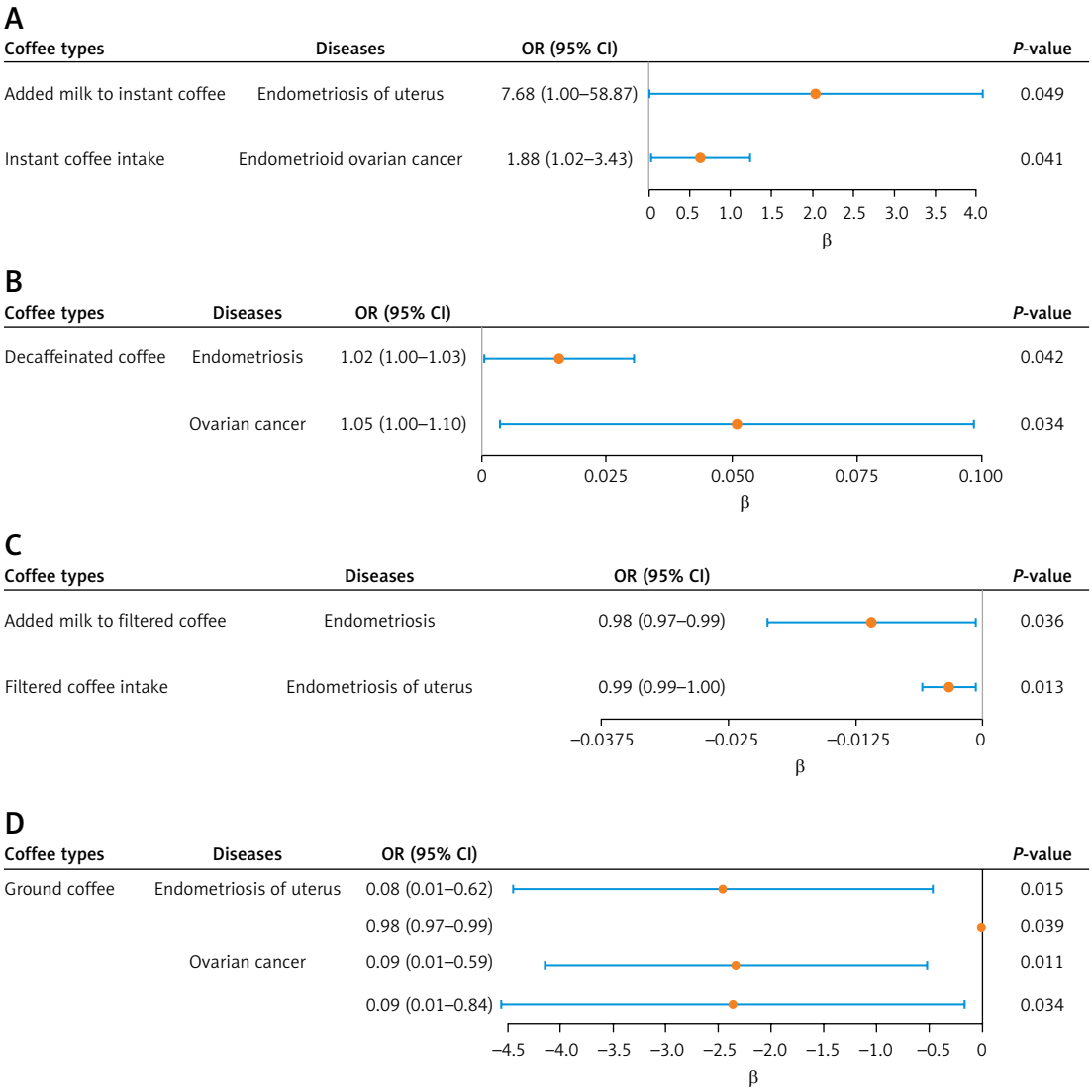

Figure 3. Forest plot of significant two-sample Mendelian randomization (2SMR) results between coffee types and gynecological diseases. Coffee types were treated as binary variables (consumed vs not consumed) and analyzed for their associations with gynecological diseases. **A** – Instant coffee and instant coffee with milk. **B** – Decaffeinated coffee. **C** – Filtered coffee and filtered coffee with milk. **D** – Ground coffee. The odds ratios (ORs) and 95% confidence intervals (CIs) are presented for each coffee type. Only significant results ($p < 0.05$) are displayed

necological diseases vary depending on the type of coffee consumed.

Evaluation of population stratification using negative control outcome

To assess the potential influence of population stratification, a negative control outcome analysis was performed using skin color as the control outcome, a trait biologically unrelated to coffee intake but potentially influenced by similar confounders. No significant associations were found between coffee intake (or its subtypes) and skin color, with *p*-values ranging from 0.399 to 0.889 (Table I). Similarly, no significant associations were observed between gynecological diseases and skin color, with *p*-values ranging from 0.081 to 0.978 (Table II). These findings indicate that residual population stratification or genetic confounding is unlikely to have biased the causal associations observed in this study.

Discussion

In this study, we conducted a two-sample Mendelian randomization (2SMR) analysis to explore the association between coffee intake, coffee types, and gynecological diseases using the IEU GWAS dataset. Our findings suggest that coffee intake is linked to a lower risk of ovarian cancer and cysts, but a higher risk of endometrioid ovarian cancer. Ground coffee appears to offer more protective benefits, particularly for ovarian cancer and endometriosis, compared to other types. Endometriosis seems more sensitive to coffee, with ground coffee, filtered coffee and filtered coffee with milk lowering its risk, although decaffeinated coffee may increase the risk of ovarian cancer and endometriosis. These results highlight the potential implications of coffee intake for gynecological health.

The Mendelian randomization design employed in this study offers key advantages over traditional observational studies, notably by reducing susceptibility to confounding. By utilizing genetic variants as instrumental variables, MR allows for more reliable causal inferences, overcoming biases associated with unmeasured confounders. Previous observational studies on coffee intake and gynecological diseases have yielded mixed results, with some suggesting protective effects, particularly for ovarian cancer, while others show no significant association or even adverse effects [11]. Such inconsistencies are likely due to confounding. MR mitigates these issues, providing stronger causal evidence [21]. Although MR studies have linked coffee consumption to improved metabolic health [12], its impact on gynecological diseases remains uncertain, necessitating further investigation into the long-term effects of different coffee types and preparation methods.

Caffeine, the primary bioactive compound in coffee, has been linked to endometrial cancer [2] and is also implicated in endometriosis. In our analysis, coffee intake was associated with an increased risk of endometrioid ovarian cancer. However, the relationship between coffee consumption and endometriosis remains inconclusive. Our data showed a minimal negative association between coffee intake and endometriosis (OR = 0.99, 95% CI: 0.99–1.00), with a narrow confidence interval indicating no statistical significance. The weak association may be due to unmeasured confounders or sample size limitations, warranting further investigation. Conversely, the results for endometrioid ovarian cancer exhibited a larger odds ratio (95% CI: 1.05–59.22), but the wide confidence intervals suggest considerable uncertainty. These findings should be interpreted cautiously, and additional studies with larger samples and refined estimates are needed to validate these associations and uncover underlying mechanisms.

In our analysis, we compared filtered coffee (without milk) with decaffeinated coffee to isolate the impact of caffeine while controlling for preparation methods. This approach minimizes confounding factors, such as milk addition. However, comparisons involving filtered coffee with milk and decaffeinated coffee introduce multiple confounders, including milk content and caffeine levels, limiting interpretability. Future studies should prioritize comparisons between coffee types prepared using identical techniques, differing primarily in caffeine content. Variations in preparation methods (e.g., filtering, brewing time) can alter coffee's chemical composition, potentially confounding the results. Controlling for these methodological differences is crucial to ensure reliable causal inferences, especially when evaluating subcategories such as filtered coffee with milk, where milk may interact with bioactive compounds in coffee. Our findings underscore the importance of carefully considering coffee type when assessing its effects on gynecological diseases, particularly endometriosis.

Coffee contains over 1000 bioactive compounds, including caffeine, chlorogenic acid, and cafestol, which may influence gynecological health. These compounds exhibit anti-inflammatory and antioxidant properties that may reduce ovarian cancer risk, with caffeine inhibiting inflammatory pathways (e.g., NF-κB signaling) and reducing oxidative stress, both implicated in ovarian carcinogenesis. Chlorogenic acid may modulate immune responses and suppress tumor progression. However, the increased risk of endometrioid

Table I. MR results for the association between coffee intake and skin color

id.exposure	id.outcome	Outcome	Exposure	Method	nsnp	b	se	pval	lo_ci	up_ci	or	or_lci95	or_uci95
ukb-b-12558	ukb-b-19560	Skin color || id:ukb-b-19560	Decaffeinated coffee || id:ukb-b-12558	Wald ratio	1	0.008	0.060	0.889	–0.110	0.126	1.008	0.896	1.135
ukb-b-12670	ukb-b-19560	Skin color || id:ukb-b-19560	Other coffee type || id:ukb-b-12670	Wald ratio	1	0.029	0.107	0.787	–0.181	0.239	1.029	0.835	1.269
ukb-b-1338	ukb-b-19560	Skin color || id:ukb-b-19560	Intake of artificial sweetener added to coffee || id:ukb-b-1338	Wald ratio	1	0.034	0.041	0.399	–0.045	0.114	1.035	0.956	1.121
ukb-b-4051	ukb-b-19560	Skin color || id:ukb-b-19560	Added milk to instant coffee || id:ukb-b-4051	Wald ratio	1	–0.021	0.040	0.601	–0.101	0.058	0.979	0.904	1.060
ukb-b-5237	ukb-b-19560	Skin color || id:ukb-b-19560	Coffee intake || id:ukb-b-5237	Inverse variance weighted	44	–0.043	0.027	0.106	–0.096	0.009	0.958	0.909	1.009
ukb-b-5332	ukb-b-19560	Skin color || id:ukb-b-19560	Added milk to filtered coffee || id:ukb-b-5332	Inverse variance weighted	2	0.027	0.046	0.553	–0.063	0.117	1.028	0.939	1.124
ukb-b-748	ukb-b-19560	Skin color || id:ukb-b-19560	Filtered coffee intake || id:ukb-b-748	Inverse variance weighted	3	–0.067	0.041	0.100	–0.147	0.013	0.935	0.863	1.013
ukb-b-930	ukb-b-19560	Skin color || id:ukb-b-19560	Instant coffee intake || id:ukb-b-930	Inverse variance weighted	2	–0.052	0.040	0.192	–0.130	0.026	0.949	0.878	1.027
ukb-b-9508	ukb-b-19560	Skin color || id:ukb-b-19560	Coffee consumed || id:ukb-b-9508	Inverse variance weighted	3	–0.189	0.099	0.055	–0.383	0.004	0.828	0.682	1.004
ukb-d-1508_1	ukb-b-19560	Skin color || id:ukb-b-19560	Coffee type: Decaffeinated coffee (any type) || id:ukb-d-1508_1	Inverse variance weighted	2	0.261	0.306	0.394	–0.338	0.860	1.298	0.713	2.362
ukb-d-1508_2	ukb-b-19560	Skin color || id:ukb-b-19560	Coffee type: Instant coffee || id:ukb-d-1508_2	Inverse variance weighted	4	0.102	0.111	0.355	–0.115	0.319	1.108	0.892	1.376
ukb-d-1508_3	ukb-b-19560	Skin color || id:ukb-b-19560	Coffee type: Ground coffee (include espresso, filter, etc.) || id:ukb-d-1508_3	Inverse variance weighted	20	–0.145	0.074	0.050	–0.289	0.000	0.865	0.749	1.000
ukb-d-1508_4	ukb-b-19560	Skin color || id:ukb-b-19560	Coffee type: Other type of coffee || id:ukb-d-1508_4	Wald ratio	1	0.007	0.647	0.991	–1.261	1.276	1.008	0.283	3.582

nsnp – the number of single nucleotide polymorphisms used as instrumental variables, b – the effect size estimate for the association between the exposure and outcome, se – the standard error of the effect size estimate, pval – the p-value for testing the null hypothesis (indicating statistical significance), lo_ci – the lower bound of the 95% confidence interval for the effect size, up_ci – the upper bound of the 95% confidence interval for the effect size, Or – the odds ratio, representing the magnitude of the association between exposure and outcome, or_lci95 – the lower bound of the 95% confidence interval for the odds ratio, or_uci95 – the upper bound of the 95% confidence interval for the odds ratio.

Table II. MR results for the association between diseases of interest and skin color

id.exposure	id.outcome	Outcome	Exposure	Method	nsnp	b	se	pval	lo_ci	up_ci	or	or_lci95	or_uci95
ieu-a-1120	ukb-b-19560	Skin color || id:ukb-b-19560	Ovarian cancer || id:ieu-a-1120	Inverse variance weighted	10	–0.030	0.017	0.081	–0.064	0.004	0.970	0.938	1.004
ieu-a-1123	ukb-b-19560	Skin color || id:ukb-b-19560	Invasive mucinous ovarian cancer || id:ieu-a-1123	Inverse variance weighted	2	0.006	0.009	0.467	–0.011	0.023	1.006	0.989	1.023
ieu-a-1229	ukb-b-19560	Skin color || id:ukb-b-19560	Serous ovarian cancer: low grade and low malignant potential || id:ieu-a-1229	Inverse variance weighted	7	0.000	0.005	0.979	–0.010	0.010	1.000	0.990	1.010
ieu-a-1230	ukb-b-19560	Skin color || id:ukb-b-19560	Serous ovarian cancer: low malignant potential || id:ieu-a-1230	Inverse variance weighted	3	0.002	0.003	0.540	–0.004	0.008	1.002	0.996	1.008
ieu-a-1231	ukb-b-19560	Skin color || id:ukb-b-19560	Mucinous ovarian cancer: invasive and low malignant potential || id:ieu-a-1231	Inverse variance weighted	4	–0.001	0.007	0.835	–0.014	0.012	0.999	0.986	1.012
ieu-a-1232	ukb-b-19560	Skin color || id:ukb-b-19560	Low malignant potential mucinous ovarian cancer || id:ieu-a-1232	Inverse variance weighted	3	0.000	0.008	0.991	–0.015	0.015	1.000	0.985	1.015
ieu-a-1233	ukb-b-19560	Skin color || id:ukb-b-19560	Low malignant potential ovarian cancer || id:ieu-a-1233	Inverse variance weighted	3	0.007	0.007	0.318	–0.007	0.022	1.008	0.993	1.022
ukb-b-10903	ukb-b-19560	Skin color || id:ukb-b-19560	Non-cancer illness code, self-reported: endometriosis || id:ukb-b-10903	Inverse variance weighted	3	–0.053	0.684	0.938	–1.394	1.288	0.948	0.248	3.626
finn-b-N14_OVARYCYST	ukb-b-19560	Skin color || id:ukb-b-19560	Ovarian cyst || id:finn-b-N14_OVARYCYST	Wald ratio	1	0.008	0.011	0.481	–0.014	0.030	1.008	0.986	1.030
ieu-a-1124	ukb-b-19560	Skin color || id:ukb-b-19560	Clear cell ovarian cancer || id:ieu-a-1124	Wald ratio	1	–0.003	0.005	0.502	–0.014	0.007	0.997	0.987	1.007
ieu-b-4963	ukb-b-19560	Skin color || id:ukb-b-19560	Ovarian cancer || id:ieu-b-4963	Wald ratio	1	1.503	0.892	0.092	–0.245	3.251	4.493	0.782	25.807
ukb-d-IBD_ENDOMETRIOSIS	ukb-b-19560	Skin color || id:ukb-b-19560	Endometriosis, IBD co-morbidity || id:ukb-d-IBD_ENDOMETRIOSIS	Wald ratio	1	1.244	1.450	0.391	–1.598	4.087	3.471	0.202	59.584
ukb-d-N80	ukb-b-19560	Skin color || id:ukb-b-19560	Diagnoses - main ICD10: N80 Endometriosis || id:ukb-d-N80	Wald ratio	1	1.244	1.450	0.391	–1.598	4.087	3.471	0.202	59.584

nsnp – the number of single nucleotide polymorphisms used as instrumental variables, b – the effect size estimate for the association between the exposure and outcome, se – the standard error of the effect size estimate, pval – the p-value for testing the null hypothesis (indicating statistical significance), lo_ci – the lower bound of the 95% confidence interval for the effect size, up_ci – the upper bound of the 95% confidence interval for the effect size, Or – the odds ratio, representing the magnitude of the association between exposure and outcome, or_lci95 – the lower bound of the 95% confidence interval for the odds ratio, or_uci95 – the upper bound of the 95% confidence interval for the odds ratio.

ovarian cancer could be linked to caffeine's effect on estrogen metabolism, as this subtype is hormonally driven. Caffeine's modulation of estrogen levels may increase estrogen bioavailability, potentially explaining the observed risk. Additionally, ovarian cancer subtypes, such as endometrioid cancer, share genetic and hormonal features with conditions such as endometriosis, which may amplify coffee's hormonal effects. These findings highlight the need for further research on the dose-dependent effects of coffee components on estrogen metabolism and their interactions with genetic and lifestyle factors.

While vertical pleiotropy, where a genetic variant affects multiple traits through the same biological pathway, poses minimal risk in MR studies, it still requires attention. MR-Egger regression revealed no significant horizontal pleiotropy, supporting the validity of the genetic instruments used. However, vertical pleiotropy could introduce subtle biases, necessitating more advanced methods in future studies to distinguish it from true causal effects. To mitigate population stratification bias, we performed a negative control analysis using skin color, which is independent of coffee intake but subject to similar confounding. No associations were found, reinforcing the validity of our causal estimates. Future MR studies should continue using such controls to enhance robustness and address unmeasured confounding.

Several limitations must be acknowledged in this study. First, Mendelian randomization assumes that genetic variants are not linked to confounders, which could introduce bias. Additionally, the lack of data on participants' age and duration of coffee exposure limits our ability to assess long-term effects. The generalizability of our findings may be affected by the specific study populations. Another limitation is the potential for reverse causation, especially with decaffeinated coffee, as individuals with health conditions may switch to decaffeinated coffee, biasing the associations. Longitudinal data are needed to better assess causality. Moreover, we did not take into account individual metabolic differences, such as variations in caffeine metabolism (e.g., CYP1A2), which could influence outcomes. Future studies should address these factors. Finally, we did not consider broader dietary habits, which may interact with coffee consumption and modify its effects. Future research should incorporate these variables for a more comprehensive understanding of coffee's impact on health.

While our study offers valuable insights, several directions for future research remain. First, further investigation is needed to elucidate the biological mechanisms underlying the observed associations, focusing on specific coffee components, their interactions, and changes during processing. Second, studying diverse populations and considering potential effect modifiers, such as genetic variations and lifestyle factors, will help assess the generalizability of our findings and enhance understanding of the interplay between coffee consumption and gynecological diseases. Lastly, long-term prospective studies and randomized controlled trials are essential to establish causality and inform evidence-based guidelines for coffee consumption in relation to gynecological health. Our findings carry significant implications for clinical practice and public health strategies. Healthcare professionals can consider the differential effects of coffee types when advising patients, offering personalized recommendations based on individual risk profiles and medical histories to optimize benefits and minimize risks. Public health initiatives could leverage these findings to design targeted interventions, promote healthier coffee choices, and raise awareness of the potential associations between coffee consumption and gynecological disease risks.

In conclusion, our study suggests that coffee consumption, particularly ground coffee, may reduce the risk of gynecological diseases such as ovarian cancer and endometriosis. However, decaffeinated coffee appears to increase these risks. Using Mendelian randomization, we emphasize the need for further research to explore the biological mechanisms and the impact of coffee type, genetic factors, and lifestyle on these associations. Future studies should guide evidence-based health strategies and clinical recommendations.

Data availability statement

Two-sample Mendelian randomization analysis was performed using the R package TwoSampleMR (v.0.5.6). The in-house R scripts used to perform 2SMR analysis and generate figures were available on GitHub (https://github.com/lynnLW/2SMR_coffee).

Acknowledgments

The authors of this study express their gratitude to all investigators involved in sharing the data.

Funding

The study is funded by the Health Commission of Henan Province (LHGJ20210766) and Zhengzhou Central Hospital Affiliated to Zhengzhou University (KYQDJJ2021002). The study design, data collection and analysis, decision to publish, and manuscript preparation were independent of any influence or involvement from sponsors or funders.

Ethical approval

Not applicable.

Conflict of interest

The authors declare no conflict of interest.

References

1. Afrin S, AlAshqar A, El Sabeh M, et al. Diet and nutrition in gynecological disorders: a focus on clinical studies. Nutrients 2021; 13: 1747.
2. Zhu G, Li Z, Tang L, et al. Associations of dietary intakes with gynecological cancers: findings from a cross-sectional study. Nutrients 2022; 14: 5026.
3. Ciebiera M, Esfandyari S, Siblini H, et al. Nutrition in gynecological diseases: current perspectives. Nutrients 2021; 13: 1178.
4. Butt MS, Sultan MT. Coffee and its consumption: benefits and risks. Crit Rev Food Sci Nutr 2011; 51: 363-73.
5. Kim Y, Je Y, Giovannucci E. Coffee consumption and all-cause and cause-specific mortality: a meta-analysis by potential modifiers. Eur J Epidemiol 2019; 34: 731-52.
6. Poole R, Kennedy OJ, Roderick P, et al. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. BMJ 2017; 359: j5024.
7. Dranoff JA. Coffee, adenosine, and the liver. Purinergic Signal 2024; 20: 21-8.
8. Muriel P, Arauz J. Coffee and liver diseases. Fitoterapia 2010; 81: 297-305.
9. Ochoa-Rosales C, der Schaft N, Braun KVE, et al. C-reactive protein partially mediates the inverse association between coffee consumption and risk of type 2 diabetes: the UK Biobank and the Rotterdam study cohorts. Clin Nutr 2023; 42: 661-9.
10. Zhao Y, Lai Y, Konijnenberg H, et al. Association of coffee consumption and prediagnostic caffeine metabolites with incident Parkinson disease in a population-based cohort. Neurology 2024; 102: e209201.
11. Baspinar B, Eskici G, Ozcelik AO. How coffee affects metabolic syndrome and its components. Food Funct 2017; 8: 2089-101.
12. Raoofi A, Rezaie MJ, Delbari A, et al. Therapeutic potentials of the caffeine in polycystic ovary syndrome in a rat model: via modulation of proinflammatory cytokines and antioxidant activity. Allergol Immunopathol (Madr) 2022; 50: 137-46.
13. Meshkani M, Saedisomeolia A, Yekaninejad M, et al. The effect of green coffee supplementation on lipid profile, glycemic indices, inflammatory biomarkers and anthropometric indices in iranian women with polycystic ovary syndrome: a randomized clinical trial. Clin Nutr Res 2022; 11: 241-54.
14. Nagata C, Kabuto M, Shimizu H. Association of coffee, green tea, and caffeine intakes with serum concentrations of estradiol and sex hormone-binding globulin in premenopausal Japanese women. Nutr Cancer 1998; 30: 21-4.
15. Kotsopoulos J, Eliassen AH, Missmer SA, et al. Relationship between caffeine intake and plasma sex hormone concentrations in premenopausal and postmenopausal women. Cancer 2009; 115: 2765-74.
16. Karaismailoglu S, Tuncer M, Bayrak S, et al. The perinatal effects of maternal caffeine intake on fetal and neonatal brain levels of testosterone, estradiol, and dihydrotestosterone in rats. Naunyn Schmiedebergs Arch Pharmacol 2017; 390: 827-38.
17. Colombo R, Papetti A. Decaffeinated coffee and its benefits on health: focus on systemic disorders. Crit Rev Food Sci Nutr 2021; 61: 2506-22.
18. Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. J Am Soc Nephrol 2016; 27: 3253-65.
19. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife 2018; 7: e34408.
20. Sanderson E, Richardson TG, Hemani G, Smith GD. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. Int J Epidemiol 2021; 50: 1350-61.
21. Lawlor DA, Tilling K, Smith GD. Triangulation in aetiological epidemiology. Int J Epidemiol 2016; 45: 1866-86.