Background: Crohn’s disease (CD) and ulcerative colitis (UC) are widely associated with smoking in epidemiological studies, whereas there are conflicting results for the association between CD and UC for both coffee and alcohol consumption. Herein, we aimed to investigate whether cigarette smoking and alcohol and coffee consumption are causally associated with either CD or UC.

Methods: We utilized 540 genome-wide significant single-nucleotide polymorphisms for 3 potentially addictive substances—nicotine, alcohol, and caffeine—to assess the association of smoking, coffee, and alcohol consumption with CD and UC (12,194 CD cases, 12,366 UC cases, and 25,042 controls of European ancestry), using Mendelian randomization analysis. Mendelian randomization estimates were used to evaluate the effect of the exposure factors on CD and UC risk. Sensitivity analysis was employed to test for any directional pleiotropy.

Results: We found evidence for a positive causal association between the age of smoking initiation and UC risk and between alcohol consumption and CD risk, which disappeared after sensitivity analysis for both associations ($P > 0.05$). No evidence for a causal association between cigarettes per day, smoking initiation, smoking cessation, and coffee consumption variables and UC or CD was found.

Conclusions: We found no clear evidence that either genetically predicted smoking, coffee consumption, or alcohol consumption are causally associated with the risk for CD or UC, although our findings indicate a potential positive association between the age of smoking and UC and between alcohol consumption and CD.

Key Words: Mendelian randomization, smoking, alcohol, coffee, inflammatory bowel diseases

INTRODUCTION

Idiopathic inflammatory bowel diseases (IBDs), which comprise Crohn’s disease (CD) and ulcerative colitis (UC), are chronic relapsing inflammatory disorders of the gastrointestinal tract.1, 2 Crohn’s disease can involve any part of the gastrointestinal tract—most commonly the terminal ileum and colon—in a noncontinuous fashion, and the inflammation is often transmural. Ulcerative colitis mostly involves the rectum and maybe part of the colon, spreading in a continuous fashion, and the inflammations are more superficial and limited to the mucosa.2, 3 The mechanism triggering IBD involves the inappropriate inflammatory response to intestinal microbes, which is mediated by the complex interplay between genetic susceptibility, environmental factors, and microbial dysbiosis.4 However, the exact mechanism behind IBD pathogenesis is not yet fully elucidated; consequently, additional research is needed to identify potentially causal risk factors that could help guide prevention efforts.

Cigarette smoking and coffee and alcohol consumption have been extensively studied for their association with inflammatory bowel disease (IBD) pathogenesis.1, 2, 5, 6 Smoking is an important modifiable factor that is consistently associated with IBD onset. Multiple observational studies have described a positive association between cigarette smoking and the risk of CD, whereas published evidence indicates that cigarette smoking has a protective role in UC pathogenesis.7 Nevertheless, former smoking is suggested to increase the risk of UC.7 It is not yet fully understood how smoking is involved in IBD risk, due to the complex composition of tobacco smoke.8 The role of alcohol consumption in the development of IBD is still controversial, with most studies investigating the role of alcohol solely in UC.5, 9, 11 There is a trend of an inverse association between coffee consumption and UC,12, 13 and yet,
these results are inconsistent in the literature, whereas no association has been observed with CD. In general, there is limited and conflicting observational evidence on the association of either coffee or alcohol consumption and risk of CD or UC. Epidemiological studies have failed to find clear associations of smoking, coffee consumption and alcohol consumption with CD or UC risk. This lack of association is probably because of the practical inability in conducting randomized controlled trials for any of the 3 modifiable risk factors to assess potential causal effect on CD or UC. Furthermore, studies investigating the role of cigarette smoking, coffee consumption, or alcohol intake on CD or UC risk are mainly observational, which are prone to biases such as reverse causation and residual confounding.

To address these limitations, we adopted a Mendelian randomization (MR) approach, which utilizes genetic variation to examine the potential causal effect of an exposure on a disease. The advantage of this approach is the random assortment of genes from parents to offspring, which renders it independent of the biases inherent in standard observational studies. Herein, we use a 2-sample MR approach to investigate whether potentially modifiable risk factors such as smoking, coffee consumption, or alcohol consumption have a causal and not a pleiotropic effect on CD and UC risk.

**METHODS**

**Selection of Genetic Variants and Data Sources**

**Exposures: smoking, alcohol consumption, and coffee consumption**

Effect estimates for exposure single-nucleotide polymorphisms (SNPs) used as instruments in the MR analysis were retrieved from a number of large-scale genome-wide association studies (GWASs).

Data for the association between the genome-wide associated SNPs regarding age of initiation of regular smoking (age of smoking, continuous variable, N = 341,427), cigarettes per day (CPD, continuous variable, N = 337,334), smoking initiation (binary variable indicating whether an individual had ever smoked regularly, N = 1,232,091), smoking cessation (binary variable contrasting current vs former smokers, N = 547,219) and Drinks per Week (N = 941,280) were obtained from a large meta-analysis of 33 GWAS studies.

Effect estimates for SNPs associated with coffee consumption were obtained from a genome-wide meta-analysis of 28 studies (N = 91,462).

For each exposure phenotype, we selected independent SNPs that achieved genome-wide significance (P = 5 x 10^-8) in each GWAS study. Pleiotropic SNPs were identified through Phenoscanner, a curate database of genotype-phenotype associations. Pleiotropic SNPs were included in the analysis, and in the case that horizontal pleiotropy was detected through sensitivity analysis, they were then excluded. The strength of the relationship of each genetic instrument (SNP) with exposure phenotype was measured with F statistic. Finally, a total of 540 independent SNPs were selected as instrumental variables for the MR analysis: 10 SNPs for age of smoking, 47 for CPD, 363 for smoking initiation, 22 for smoking cessation, 90 for drinks per week, and 8 for coffee consumption (Supplementary Tables 1–6).

**Outcomes: Crohn’s disease and ulcerative colitis**

Summary data for the associations between the selected SNPs and the 2 types of IBD (ie, CD and UC) were obtained from the largest GWAS meta-analysis published by de Lange et al. The study included British participants of European ancestry from UK Inflammatory Bowel Disease Genetics (UKIBDGC) and UK10K consortia (CD cases, N = 12,194; UC cases, N = 12,366; controls, N = 25,042). DNA was extracted from the blood or saliva of IBD cases and controls. Nine million variants were tested for their association with CD and UC after genotyping with Human Core Exome and imputation using the 1000 Genomes reference panel (Phase 3v5).

**Mendelian Randomization Analysis**

We used summary genetic associations from 2 different GWAS to test each association through 2-sample MR design. In the main analyses, 2 conventional MR methods were used: the random-effects inverse variance weighted (IVW) and the maximum likelihood. The advantage of a random-effects model is that it accounts for the difference between the effect sizes of the selected SNPs for the exposure phenotype. The IVW method is based on a meta-analysis of the Wald ratio, which is the ratio of the estimates of the 2 genetic associations (ie, SNP-exposure over SNP-outcome estimate). The maximum likelihood is an alternative approach that estimates the causal effect by direct maximization of the likelihood given the SNP-exposure and SNP-outcome effects and assuming a linear relationship between the exposure and outcome. Mendelian randomization analysis between coffee consumption and the 2 types of IBD was performed twice: (1) using all 8 genome-wide associated SNPs that resulted from a transethnic meta-analysis that included participants of European ancestry and a follow-up meta-analysis that included participants of mixed ancestry, and (2) using 4 SNPs (rs17685, rs2472297, rs4410790, rs7800944) derived from a meta-analysis of European ancestry only participants. To address multiple testing, a Bonferroni-corrected P value of 0.004 (ie, 0.05/12 putative risk factors) was considered significant.

Statistical power for the MR analysis on CD and UC was calculated using an online power calculator (http://cnsenomics.com/shiny/mRnd/). The SNPs for age of smoking, CPD,
smoking cessation, coffee consumption, and alcohol consumption explained an estimated 0.1%, 1.4%, 0.2%, 1.4%, 0.3%, and 0.5% of phenotypic variability, respectively. Given a type 1 error of 5%, using data by Liu et al. and Cornellis et al., we had 80% power to detect a causal association of smoking, coffee consumption, and alcohol consumption on CD or UC, when the expected odds ratios (OR) per 1 SD were ≤0.75, 0.55, and 0.68, respectively (Supplementary Table 8).

Sensitivity Analysis

For the MR analyses to be valid, 3 important assumptions should hold: (1) a true association is required between the genetic instruments and the exposure of interest each time, (2) the genetic markers should affect CD or UC only through an alternative causal pathway, and (3) the genetic markers are independent of any confounders between the exposure of interest and the disease. If the second assumption is violated and SNPs influence the risk of the outcome through pleiotropic pathways, then the MR results are not valid. To assess for any horizontal pleiotropy (ie, one genetic variant has independent effects on multiple traits, not just the outcome being examined) or bias, sensitivity analyses such as MR-Egger and weighted median estimator were performed. The MR-Egger approach exploits the Egger regression tool, which is used to detect small study bias in meta-analysis; but in the case of MR analysis, it is adapted to assess whether genetic variants have pleiotropic effects on the outcome that differ on average from zero (directional pleiotropy). The MR-Egger slope provides an estimate of the pleiotropy-corrected causal estimate, but this is underpowered if the proportion of the variance in the exposure explained by the SNPs is small. The weighted median approach is complementary to the MR-Egger regression method, and its estimate is valid as long as at least 50% of the genetic variants (SNPs) are not pleiotropic. In the case of pleiotropy, the evaluation was based on the intercept obtained from MR-Egger analysis.

Heterogeneity between genetic variants was estimated using the Cochran Q test. In case there was evidence for heterogeneity, a “leave-one-out” sensitivity analysis was performed to test whether a single genetic variant was driving the causal association.

All statistical analyses were performed using the Mendelian Randomization R package.

Ethical Considerations

All participants in the UKIBDGC and UK10K studies gave their informed consent to participation.

RESULTS

Selection of Genetic Variants for Mendelian Randomization

Overall, 540 Linkage Disequilibrium-independent SNPs that were genome-wide associated with the 3 risk factors were selected. Table 1 shows the sample size of a GWAS meta-analysis for CD and UC used in the current study for the extraction of summary results for those 540 SNPs. Figure 1 shows a schematic representation of the current MR analysis and the number of SNPs selected for each exposure phenotype. Summary data for the estimates of the associations of the genetic variants with smoking, coffee consumption, and alcohol consumption respectively are presented in Supplementary Tables 1–6. The estimates of the associations of the genetic variants with the risk for CD and UC are also presented in Supplementary Tables 1–6. These SNPs were all genome-wide associated with each exposure phenotype, but none was genome-wide associated with any IBD type risk at a genome-wide level. The selected SNPs are relatively strong instruments for the corresponding addictive risk factor, with F statistic values greater than the recommended threshold of 10—with the exception of one SNP (rs9902453) associated with coffee consumption, which has an F statistic value of 9.

Mendelian Randomization Analysis

Figures 2 and 3 show scatter plots of associations between smoking, coffee consumption, and alcohol consumption polymorphisms and risk of CD and UC, respectively. Mendelian randomization results supported a suggestive causal effect of the age of initiation of regular smoking on UC risk (OR, 2.38; 95% confidence interval [CI], 1.13–5.02; \( P = 0.02 \)). Neither the IVW nor the maximum likelihood method revealed any evidence to support a causal association between multi-SNP risk scores for the
remaining smoking phenotypes and UC risk or between any of the smoking phenotypes and CD risk (Table 2). The overall estimates calculated by maximum likelihood and IVW approaches did not reveal any association between coffee consumption and any of the 2 IBD subtypes (CD: OR, 1.29, 95% CI, 0.83–2.00, \( P = 0.26 \); UC: OR, 0.96, 95% CI, 0.75–1.23, \( P = 0.78 \)). There was also a suggestive positive association detected between alcohol consumption and CD risk (OR, 1.73, 95% CI, 1.04–2.86, \( P = 0.03 \)) but not between alcohol consumption and UC risk (OR, 1.57, 95% CI, 0.92–2.66, \( P = 0.10 \)). Mendelian randomization analysis showed that 1-SD increase in alcohol consumption (drinks/week) was causally associated with a 73% relative increase in Crohn's disease risk (OR, 1.73; 95% CI, 1.04–2.86; \( P = 0.03 \)).

**Sensitivity Analysis**

There was evidence for substantial heterogeneity observed for the effect estimates of several exposures (CPD, smoking initiation, coffee consumption, and alcohol consumption) for both outcomes (CD and UC; \( F \) up to 96%). This heterogeneity was confirmed for some exposure phenotypes, with the weighted median analysis yielding opposite results when compared with the IVW approach.

There was evidence of horizontal pleiotropy (the SNPs affect the 2 traits through independent pathways) in the analysis of CPD and CD based on the result of the intercept term of the MR-Egger regression (\( P = 0.02 \); Supplementary Table 7). Although, the leave-one-out analysis indicated that horizontal pleiotropy did not significantly modify the MR results. Although there was no heterogeneity detected for the age of smoking SNPs
and UC, the MR-Egger results were not consistent with those generated by the IVW and the maximum likelihood (Supplementary Table 7). Due to the unlikely large effect estimate (>2-fold increase), we proceeded with the leave-one-out analysis, which showed that when each of the 3 SNPs—rs1403174, rs11915747, or rs12611472—was excluded, the estimated effect size decreased, showing weak evidence for a causal association.

Sensitivity analysis (MR-Egger multivariable analysis) did not reveal any horizontal pleiotropy for the estimates of genetic variants for alcohol consumption on CD risk. However, we proceeded with the leave-one-out analysis due to the high detected heterogeneity ($I^2 = 96\%$) and the strong association between one of the genetic instruments (rs1260326) and CD, although not reaching genome-wide

<table>
<thead>
<tr>
<th>Exposure Phenotype</th>
<th>Number of GIs</th>
<th>IVW OR (95% CI); $P$-value</th>
<th>Maximum Likelihood OR (95% CI); $P$-value</th>
<th>Heterogeneity $I^2$ ($P$-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crohn’s Disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of Smoking</td>
<td>10</td>
<td>1.52 (0.66–3.52); 0.33</td>
<td>1.55 (0.66–3.64); 0.32</td>
<td>0% (0.12)</td>
</tr>
<tr>
<td>CPD</td>
<td>47</td>
<td>0.89 (0.65–1.20); 0.44</td>
<td>0.95 (0.65–1.20); 0.44</td>
<td>96% (&lt;0.001)</td>
</tr>
<tr>
<td>Smoking Initiation</td>
<td>363</td>
<td>0.87 (0.77–1.00); 0.06</td>
<td>0.87 (0.76–1.00); 0.055</td>
<td>40% (&lt;0.001)</td>
</tr>
<tr>
<td>Smoking Cessation</td>
<td>22</td>
<td>1.19 (0.84–1.70); 0.33</td>
<td>1.19 (0.83–1.71); 0.34</td>
<td>79% (0.001)</td>
</tr>
<tr>
<td>Coffee Consumption</td>
<td>8</td>
<td>1.29 (0.83–2.00); 0.26</td>
<td>1.34 (0.84–2.13); 0.23</td>
<td>93% (&lt;0.001)</td>
</tr>
<tr>
<td>Coffee Consumption</td>
<td>4</td>
<td>1.13 (0.84–1.52); 0.42</td>
<td>1.13 (0.84–1.52); 0.41</td>
<td>93% (0.3)</td>
</tr>
<tr>
<td>Alcohol Consumption</td>
<td>90</td>
<td>1.73 (1.04–2.86); <strong>0.03</strong></td>
<td>1.77 (1.06–2.97); <strong>0.03</strong></td>
<td>94% (&lt;0.001)</td>
</tr>
<tr>
<td><strong>Ulcerative Colitis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of Smoking</td>
<td>10</td>
<td>2.38 (1.13–5.02); <strong>0.02</strong></td>
<td>2.45 (1.15–5.22); <strong>0.02</strong></td>
<td>0% (0.16)</td>
</tr>
<tr>
<td>CPD</td>
<td>47</td>
<td>0.95 (0.70–1.30); 0.77</td>
<td>0.96 (0.70–1.31); 0.77</td>
<td>96% (&lt;0.001)</td>
</tr>
<tr>
<td>Smoking Initiation</td>
<td>363</td>
<td>0.90 (0.79–1.03); 0.13</td>
<td>0.90 (0.79–1.03); 0.14</td>
<td>41% (&lt;0.001)</td>
</tr>
<tr>
<td>Smoking Cessation</td>
<td>22</td>
<td>1.11 (0.82–1.51); 0.49</td>
<td>1.12 (0.82–1.53); 0.47</td>
<td>79% (0.02)</td>
</tr>
<tr>
<td>Coffee Consumption</td>
<td>8</td>
<td>0.96 (0.75–1.23); 0.78</td>
<td>0.96 (0.75–1.24); 0.77</td>
<td>93% (0.08)</td>
</tr>
<tr>
<td>Coffee Consumption</td>
<td>4</td>
<td>0.92 (0.76–1.12); 0.40</td>
<td>0.92 (0.76–1.12); 0.40</td>
<td>93% (0.2)</td>
</tr>
<tr>
<td>Alcohol Consumption</td>
<td>90</td>
<td>1.57 (0.92–2.66); 0.10</td>
<td>1.61 (0.93–2.78); 0.09</td>
<td>94% (&lt;0.001)</td>
</tr>
</tbody>
</table>

Smoking initiation and smoking cessation are binary phenotypes; CPD is measured as cigarettes per day; age of smoking is measured as the age of initiation of regular smoking, coffee consumption is measured as cups of coffee/day and alcohol consumption as drinks/week. Abbreviations: OR, odds ratio representing the odds of the outcome variable with every unit increase of the exposure variable. 95% CI, lower and upper limits for a 95% confidence interval.
level of significance ($P = 2.7 \times 10^{-5}$), and hence, driving the results of the MR analysis. After the exclusion of rs1260326, the association faded away (OR, 1.48; 95% CI, 0.91–2.39; $P = 0.12$). The pleiotropic effect of rs1260326 was also confirmed through Phenoscanner.

**DISCUSSION**

Using summary data from large GWAS studies, our MR study demonstrated a possible increase in the risk of UC onset for later genetically predicted smoking initiation. There was suggestive evidence for a positive association between increased alcohol consumption and CD risk. Both associations disappeared after sensitivity analysis and the exclusion of pleiotropic variants. Additionally, there was no clear evidence for a causal effect of coffee consumption and most of the smoking phenotypes (CPD, smoking initiation, and smoking cessation) on either CD or UC risk.

The effect of smoking on the pathogenesis of IBD is comprehensively studied and the most clearly defined environmental risk factor for the development of CD and UC. Our findings provided limited support for a positive causal association between the age of smoking initiation and UC risk by the maximum likelihood and the IVW analyses. At the moment, there is no observational study showing an association between the age of smoking particularly and UC risk. A functional study showed that an earlier age of initiation of regular smoking cessation was shown to be protective against UC development, most likely through the alteration of intestinal microbiota, but the underlying mechanism is not fully elucidated yet. It remains to be studied whether the age of initiation of regular smoking or smoking alone is mostly affecting the risk for UC development.

Evidence from meta-analysis of observational studies argue that current smoking has a protective effect on UC risk, whereas current smoking increases the risk for CD. However, our MR estimates provided consistent results about smoking initiation and its association with CD and UC risk, which was close to null. Regarding smoking heaviness, there is no meta-analysis examining its effect on IBD onset, most likely due to the vagueness of the reports mentioning light, moderate, and heavy smoking, making it impossible to combine the results and examine a dose-response relationship. Herein, we examined whether there is a causal association between smoking heaviness (CPD) and CD or UC, but there was no evidence for such an association; hence, no dose-response relationship was detected.

A gene-environment interaction study identified 64 SNPs that possibly modify the effect of smoking behavior (current, ever, and never smoking) on IBD risk. Approximately, 45% of the SNPs that interact with smoking were in close vicinity to SNPs previously associated with IBD. This interaction could explain part of the reported association between smoking behavior and IBD risk in observational studies and also the present nonsignificant findings. In addition, the association between cigarette smoking and IBD risk could possibly be affected by epigenetic modifications.

This study sheds light on the conflicting results regarding the association between coffee consumption and CD and UC risk, in addition to alcohol consumption and CD and UC risk. Those conflicting results most likely arise from the lack of data on the exposure of addiction risk factors before the onset of IBD and, therefore, the lack of large studies investigating those associations. In addition, the inconsistent results may be due to study-specific biases and differential adjusting or even nonadjusting for confounding factors. There are no studies investigating the association between alcohol consumption and CD. A meta-analysis investigating the role of alcohol on the pathogenesis of IBD identified only 2 studies that successfully fulfilled their inclusive criteria and were investigating the association between alcohol intake and UC risk. However, due to the large heterogeneity between the 2 studies, it was not possible to combine the results. Although there was a protective effect for light alcohol consumers against the development of UC, this association disappeared when it was adjusted for smoking, pointing that this association is confounded by smoking. The most recent and largest meta-analysis that investigated alcohol consumption and risk for UC did not reveal any association between alcohol consumption and UC risk. These results are consistent with the findings of the present study, pointing to no causal association between alcohol intake and UC risk.

The same meta-analysis regarding alcoholic beverage consumption and risk for UC suggested a potential, but not robust, role of coffee consumption in the development of UC, after adjusting for smoking. This is again consistent with our MR results, though it is conflicting to the findings of a functional study on mice showing that acute colitis is reduced upon caffeine treatment through the downregulation of CHI3L1 expression bacterial interaction effect. Vitaly, however, the mechanism that drives the pathogenesis of IBD is complex and most likely different than the one underlying the progression of the disease, and the MR study presented here is designed to investigate only the former.

**Strengths and Limitations of the Study**

The main strength of an MR study is that it is based on the law of independent assortment, thus overcoming potential biases due to confounding and reverse causation. Reverse causation is a concern in the case of IBD, where the time-interval between the onset of the disease and the time of the diagnosis is 10 years or more. This MR study assessed multiple potentially modifiable risk factors in relation to IBD and used summary statistics from large data sets for addiction risk factors and the 2 IBD subtypes. For most of the exposure phenotypes, the genetic variants explain a sufficient proportion of the variance, with the exception of coffee consumption, where we cannot exclude that our findings on caffeine consumption might have been affected by weak instrument bias, which depends on the strength of the genetic instrument. As the investigations were undertaken in a 2-sample setting, any bias from weak instruments is usually in the direction of the null.
However, we cannot exclude the possibility that the effect size (more than a 2-fold increase of UC risk for each year of delayed smoking initiation) could be inflated due to the weak standard error. Another more likely scenario for the failure to detect a causal association between the 3 exposure phenotypes and the 2 subtypes of IBD could be the limited power of the study to detect associations of small effect sizes. Sensitivity analyses were used to look for potential violations of the assumptions. Indeed, there was evidence for pleiotropy observed, and thus the analyses were performed again with the pleiotropic variants excluded. A limitation of this 2-sample MR study was that we could not explore potential non-linear relationships between potentially modifiable exposures and disease outcomes, which could be the case for any of the 3 risk factors. For example, caffeine is protective for many diseases at a low dose, but it is harmful at high doses.

CONCLUSION

To our knowledge, this is the first study using MR analyses to address the potentially causal relationship between addictive substances and CD or UC. In conclusion, using data from large genetic consortia, we demonstrated that there is no evidence for a causal association between coffee consumption and risk of CD or UC and only indicative evidence for a positive causal association between genetically predicted age of smoking initiation and UC risk and between genetically predicted alcohol consumption and CD risk. Yet, the negative health impacts of alcohol consumption and initiation of smoking at a young age are likely to be striking and should be taken into account.

SUPPLEMENTARY DATA

Supplementary data is available at Inflammatory Bowel Diseases online.

ACKNOWLEDGMENTS

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization. The authors thank the staff and participants of the UKIBDGC and UK10K Consortiums.

REFERENCES