

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/322512586>

Investigating causal associations between use of nicotine, alcohol, caffeine, and cannabis: A two-sample bidirectional Mendelian randomization study

Article in *Addiction* · January 2018

DOI: 10.1111/add.14154

CITATIONS

17

READS

211

3 authors:



Karin Verweij

Queensland Institute of Medical Research

162 PUBLICATIONS 3,597 CITATIONS

[SEE PROFILE](#)



Jorien L. Treur

University of Bristol

63 PUBLICATIONS 1,179 CITATIONS

[SEE PROFILE](#)



Jacqueline M Vink

Radboud University

289 PUBLICATIONS 11,898 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Biological mechanisms of creativity [View project](#)



The Consortium for Causal Analysis Research in Tobacco and Alcohol (CARTA) [View project](#)

Short Report

Investigating causal associations between use of nicotine, alcohol, caffeine, and cannabis: A two-sample bidirectional Mendelian randomization study

Karin J.H. Verweij^{1*}, Jorien L. Treur^{1*}, & Jacqueline M. Vink¹

Affiliations:

1. Behavioural Science Institute, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

*Joint first authors

Running head: MR study of substance use

Word count (excluding abstract, references, tables, and figures): 2000

Declarations of competing interest: None

Keywords: Alcohol, cannabis, nicotine, caffeine, causality, Mendelian randomization, genes, gateway, pleiotropy

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/add.14154

Abstract

Background and Aims: Epidemiological studies consistently show co-occurrence of use of different addictive substances. Whether these associations are causal or due to overlapping underlying influences remains an important question in addiction research. Methodological advances have made it possible to use published genetic associations to infer causal relationships between phenotypes. In this exploratory study, we used Mendelian randomization (MR) to examine the causality of well-established associations between nicotine, alcohol, caffeine, and cannabis use.

Methods: Two-sample MR was employed to estimate bi-directional causal effects between four addictive substances: nicotine (smoking initiation and cigarettes smoked per day), caffeine (cups of coffee per day), alcohol (units per week), and cannabis (initiation). Based on existing genome-wide association results we selected genetic variants associated with the exposure measure as an instrument to estimate causal effects. Where possible we applied sensitivity analyses (MR-Egger and weighted median) more robust to horizontal pleiotropy.

Results: Most MR tests did not reveal causal associations. There was some weak evidence for a causal positive effect of genetically instrumented alcohol use on smoking initiation and of cigarettes per day on caffeine use, but these did not hold up with the sensitivity analyses. There was also some suggestive evidence for a positive effect of alcohol use on caffeine use (only with MR-Egger) and smoking initiation on cannabis initiation (only with weighted median). None of the suggestive causal associations survived corrections for multiple testing.

Conclusions: Two-sample Mendelian randomization analyses found little evidence for causal relationships between nicotine, alcohol, caffeine, and cannabis use.

Introduction

Epidemiological studies consistently show correlations between use of nicotine, alcohol, caffeine, and cannabis, such that individuals that use one substance are more likely to also use another [e.g. 1]. Two main mechanisms may explain this co-occurrence: (1) a causal relationship and (2) common liability. A *causal* relationship entails that use of one drug increases the probability of use of others. A prominent theory in line with a causal explanation is the gateway hypothesis, posing that substance use progresses in a stage-like sequence, in which use of legal substances such as alcohol and tobacco is typically followed by use of cannabis, which proceeds use of other (more dangerous) illicit drugs [2-4]. However, the sequential order of use is less clear in other cases; e.g. for the co-occurrence of smoking and alcohol use or caffeine consumption. Also, several studies found evidence supporting a reverse-gateway hypothesis (i.e. reverse causation), in which use of for example cannabis or illicit substances increases the use of nicotine or alcohol use [5,6]. Hence, the gateway hypothesis cannot explain all observed associations. Different mechanisms could underlie causal relationships. They can be due to biological effects, such as cross-sensitisation and activation of the reward system, or because use of one substance influences the metabolism or enhances the effect of another substance [7-12]. Also, social/environmental effects could play a role, for example if use of one substance causes the user to be in different social settings leading to use of other substances [13].

An alternative explanation of the co-occurrence of use of different substances is the *common liability* model, which proposes it is due to common underlying genetic and/or environmental influences [14-16]. Twin studies have shown that correlations between (ab)use of different substances can be attributed to both overlapping genetic and overlapping environmental influences [1,17]. More recently, studies using measured genetic variants have also shown that common genetic factors play a role in the co-occurrence. Vink et al. [18] found that polygenic scores for cigarettes smoked per day (CPD) predicted alcohol consumption and cannabis use. Nivard et al. [19] used genome-wide association [GWA; 20] summary data for smoking,

alcohol, cannabis, and caffeine use and showed there are substantial genetic correlations between them.

Albeit previous studies provide a clear pattern of phenotypic, genetic, and environmental correlations in the co-occurrence of substance use, the exact nature of these associations remains unclear. Genetic associations can be the result of biological, or horizontal, pleiotropy, where the same genes influence vulnerability for multiple phenotypes. But genetic correlations can also arise from causal relationships between phenotypes (mediated/vertical pleiotropy). For example, if smoking cigarettes causally increases the likelihood that someone also starts using cannabis, genes that underlie vulnerability for cigarette smoking will indirectly also be associated with cannabis use. Similarly, environmental correlations can be due to environmental confounders that increase vulnerability to use both substances, but can also reflect a causal relationship.

Whether relationships between use of different substances are causal or not remains an important question in addiction research. Because of methodological advances in the field of genetics, it is now possible to better distinguish between pleiotropic versus causal associations. In this explorative study we will use two-sample Mendelian randomization (MR) to estimate the causality of well-established associations between use of different addictive substances (nicotine, alcohol, caffeine, and cannabis).

Methods

MR is a technique that utilizes (a) genetic variant(s) strongly associated with an exposure variable, as 'instrument', to estimate causal effects on an outcome variable. This approach minimizes distorting effects of confounders and reverse causation, provided that the following assumptions are met: 1) the genetic instrument is predictive of the exposure variable, 2) the genetic instrument is independent of confounders, 3) the genetic instrument does not affect the outcome through an independent pathway, other than by its possible causal effect through the exposure (exclusion restriction assumption). Presence of horizontal pleiotropy, where the same genetic variants

influence vulnerability for multiple phenotypes, could lead to violation of assumption 2 and 3 [21].

While traditional MR requires individual-level genotype and phenotype data in a single sample, two-sample MR merely requires summary statistics of two GWA studies [22].

Bi-directional causal effects were estimated between four commonly used addictive substances (see Supplementary Figure 1); nicotine [smoking initiation (N=74,035) and CPD (N=38,181); 23], caffeine [cups of coffee per day (N=92,501); 20], alcohol [units of alcohol per week (N=112,117); 24] and cannabis [initiation (N=32,330); 25]. From these studies we identified single nucleotide polymorphisms (SNPs) that, based on the robustness of their association with the exposure variable, could serve as instrument (*gene-exposure association*). Next, SNPs were pruned for linkage disequilibrium ($r^2 < 0.01$) and remaining variants (or proxies, $r^2 \geq 0.8$) were identified in GWA summary-level data of the outcome variable (*gene-outcome association*). According to the basic principle of MR, evidence for both a gene-exposure and a gene-outcome association, provided the MR assumptions are met, would point to a causal effect. For a detailed description of two-sample MR, see [22].

The exact analysis depended on the number of SNPs used as genetic instrument. In case of a single SNP, the Wald ratio method was applied which is the gene-outcome association divided by the gene-exposure association [26]. When the instrument comprised multiple SNPs, inverse-variance weighted (IVW) linear regression was applied [27]. For genetic instruments that contained sufficient SNPs (≥ 10) we applied two sensitivity analyses more robust to horizontal pleiotropy: 1) the weighted median approach, which provides a consistent estimate of the causal effect even when up to 50% of the weight comes from invalid instruments [28] and 2) MR-Egger, which provides a consistent estimate of the causal effect as long as the strength of the genetic instrument does not correlate with the effect the instrument has on the outcome. This is known as the InSIDE assumption (Instrument Strength Independent of Direct Effect) and is considered a weaker version of the exclusion restriction assumption [29,30] Like all instrumental variable methods, MR is vulnerable to weak instrument bias. For the IVW approach, instrument bias was quantified with the F-statistic

($F > 10$ indicates the instrument is sufficiently strong [31]). We also report the I^2 statistic which quantifies heterogeneity between genetic variants in an instrument and indicates whether the 'NO Measurement Error' (NOME) assumption has been violated; if $I^2 < 0.9$ the NOME assumption is likely to be violated [32]. We also report estimates from MR-Egger SIMEX (simulation extrapolation) which adjusts for bias caused by violation of the NOME assumption [32].

Two sets of analyses were performed; one where only SNPs were included as instrument that exceeded the p-value threshold for genome-wide significance ($< 5 \times 10^{-8}$) and one with SNPs that exceeded a more lenient p-value threshold of $< 1 \times 10^{-5}$. As we tested causality of 18 associations, we adopted an alpha-level of 2.8×10^{-3} ($0.05/18$).

Results

Results for IVW analyses are shown in Table 1 and results for all sensitivity analyses (weighted median/MR-Egger/MR-Egger SIMEX) in Supplementary Table 1. Forest plots of the MR results per SNP and leave-one-out IVW analyses are presented in Supplementary Figures 2 and 3. We found no consistent evidence for causal effects, only a few suggestive findings. The IVW method pointed to a causal, positive effect of genetically instrumented alcohol use on smoking initiation and of CPD on caffeine use. However, these associations were only found for the p-value threshold of $< 1 \times 10^{-5}$ and the weighted median or MR-Egger methods did not show similar evidence (Supplementary Table 1), indicating these results are not robust to the 'no pleiotropy' assumption.

Furthermore, we found some suggestive evidence with MR-Egger regression that alcohol use increases caffeine use, but this effect was only found for one p-value cut-off and not seen with the IVW and weighted median tests. Finally, a positive impact of smoking initiation on the odds of cannabis use was found with the weighted median analysis, but not the other tests. Note that none of the suggestive causal relationships survived corrections for multiple testing.

Instrument strength was sufficient with the F-statistic being >10 for all genetic variants (see Supplementary Table 2). Cochran's heterogeneity statistic (Q), indicated heterogeneity for IVW analyses of alcohol with caffeine and cannabis use and of ever smoking with alcohol and caffeine use (see Supplementary Table 3). The MR-Egger intercepts showed some evidence ($p=0.04$) for pleiotropy for the analyses from alcohol to caffeine use (Supplementary Table 4).

Discussion

In this explorative two-sample MR study, there was very little evidence for causal relationships between the most commonly used addictive substances; nicotine, alcohol, caffeine, and cannabis. We found some indications for causal relationships ($p<0.05$), but none of these findings were consistent over the different MR tests and different p-value thresholds for selecting the genetic instruments. Moreover, none of the associations remained significant after correcting for multiple testing. Overall, these findings do not support the hypothesis that causal relationships (including the gateway model) explain the co-occurrence of use of different substances, but they could be in line with a common liability model, which poses that the co-occurrence is due to common underlying genetic and/or environmental influences.

Some relationships studied here have been investigated with MR studies before. Evidence for a causal effect of smoking heaviness on coffee consumption has been reported with a large one-sample MR study [33]. Although our results were suggestive of a positive effect, evidence was weak. It is important to note that with our approach -using summary-level data- we were unable to stratify our analyses on smoking status. Therefore, for the analyses with CPD as the exposure, we may have been underpowered to pick up causal effects because the outcome GWA study included a combination of smokers and non-smokers.

In the other direction, a recent comprehensive three-part study reported some evidence for a causal, negative effect of caffeine use on CPD with a two-sample MR method on summary-level data, but not with MR on individual-level data or in vitro experiments [34]. With a slightly different

genetic instrument, we also found mostly negative effect sizes of caffeine use on CPD, but these were not significant. In another small-scale MR study (N=180), Irons et al. [35] took a genetic variant strongly associated with alcohol use in Asians populations and, similar to us, found little evidence for alcohol as a gateway to other substance use.

Our study comes with some limitations. Two-sample MR heavily depends on the robustness of the instrumental variables. For some of our phenotypes, the genetic variants explain very little variance which may make them inadequate as instruments [25]. As larger samples become available, accuracy of SNP effects will increase and two-sample MR approaches will become more powerful. Secondly, we took an explorative approach, selecting genetic instruments from published GWA studies. Identifying genetic instruments with a clear (biological) may result in higher quality instruments, but would also be more challenging as the exact role of genetic variants associated with substance use is often unknown. For caffeine consumption our genetic instrument may have been weak when samples of the outcome variables included many individuals that do not consume coffee. This is unlikely to be a problem however, given that a genetic risk score predicted coffee consumption in a combined sample of coffee and non-coffee drinkers [36]. Finally, sample overlap between the GWA studies can cause bias in the direction of the null and may result in false-negative findings [37]. Between most studies, sample overlap was minimal (0%-5%, Supplementary Table 5), which is not likely to have affected our results [37]. However, sample overlap was considerable between caffeine use and smoking initiation (33%) and CPD (26%).

In conclusion, a previous study [19] found substantial (genome-wide) genetic correlations between use of different substances without differentiating between causal versus pleiotropic effects. Here, we performed the first two-sample MR study to estimate causality in the relationships between use of nicotine, alcohol, caffeine, and cannabis. We found little evidence for causal effects between use of these substances, which may suggest that the phenotypic associations between use of different substances are more likely to be explained by other mechanisms, such as common

liability. However, in light of our limitations, future studies with more powerful instrumental variables should be performed to further investigate the nature of the associations.

Acknowledgements

The work described here was supported by the European Research Council [grant number 284167: 'Beyond the Genetics of Addiction', principal investigator JMV]. JLT is supported by a Rubicon grant from the Netherlands Organization for Scientific Research (NWO; grant number 446-16-009).

References

1. Kendler, K.S., et al., *Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood*. Archives of General Psychiatry, 2008. **65**(6): p. 674-682.
2. Kandel, D.B., *Stages in adolescent involvement in drug-use*. Science, 1975. **190**(4217): p. 912-914.
3. Kandel, D.B., K. Yamaguchi, and L.C. Klein, *Testing the Gateway Hypothesis*. Addiction, 2006. **101**(4): p. 470-2; discussion 474-6.
4. Kandel, D. and E. Kandel, *The Gateway Hypothesis of substance abuse: developmental, biological and societal perspectives*. Acta Paediatr, 2015. **104**(2): p. 130-7.
5. Becker, J., et al., *Cannabis use and other predictors of the onset of daily cigarette use in young men: what matters most? Results from a longitudinal study*. BMC Public Health, 2015. **15**: p. 843.
6. Patton, G.C., et al., *Reverse gateways? Frequent cannabis use as a predictor of tobacco initiation and nicotine dependence*. Addiction, 2005. **100**(10): p. 1518-25.
7. Doyon, W.M., et al., *Nicotine decreases ethanol-induced dopamine signaling and increases self-administration via stress hormones*. Neuron, 2013. **79**(3): p. 530-40.
8. Benowitz, N.L., M. Peng, and P. Jacob, 3rd, *Effects of cigarette smoking and carbon monoxide on chlorzoxazone and caffeine metabolism*. Clin Pharmacol Ther, 2003. **74**(5): p. 468-74.
9. Temple, J.L., *Caffeine use in children: what we know, what we have left to learn, and why we should worry*. Neurosci Biobehav Rev, 2009. **33**(6): p. 793-806.
10. Robinson, T.E. and K.C. Berridge, *The psychology and neurobiology of addiction: an incentive-sensitization view*. Addiction, 2000. **95 Suppl 2**: p. S91-117.
11. Robinson, T.E. and K.C. Berridge, *Incentive-sensitization and addiction*. Addiction, 2001. **96**(1): p. 103-14.
12. Kenny, P.J., *Brain reward systems and compulsive drug use*. Trends Pharmacol Sci, 2007. **28**(3): p. 135-41.
13. Wagner, F.A. and J.C. Anthony, *Into the world of illegal drug use: exposure opportunity and other mechanisms linking the use of alcohol, tobacco, marijuana, and cocaine*. Am J Epidemiol, 2002. **155**(10): p. 918-25.
14. Morral, A.R., D.F. McCaffrey, and S.M. Paddock, *Reassessing the marijuana gateway effect*. Addiction, 2002. **97**(12): p. 1493-1504.

15. Palmer, R.H., et al., *Genetic etiology of the common liability to drug dependence: evidence of common and specific mechanisms for DSM-IV dependence symptoms*. Drug Alcohol Depend, 2012. **123 Suppl 1**: p. S24-32.
16. Vanyukov, M.M., et al., *Common liability to addiction and "gateway hypothesis": theoretical, empirical and evolutionary perspective*. Drug Alcohol Depend, 2012. **123 Suppl 1**: p. S3-17.
17. Ystrom, E., et al., *Genetic and environmental risk factors for illicit substance use and use disorders: Joint analysis of self and co-twin ratings*. Behav Genet, 2014. **44**(1): p. 1-13.
18. Vink, J.M., et al., *Polygenic risk scores for smoking: predictors for alcohol and cannabis use?* Addiction, 2014. **109**(7): p. 1141-51.
19. Nivard, M.G., et al., *Connecting the dots, genome-wide association studies in substance use*. Mol Psychiatry, 2016. **21**(6): p. 733-5.
20. Cornelis, M.C., et al., *Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption*. Mol Psychiatry, 2015. **20**(5): p. 647-56.
21. Davey Smith, G. and G. Hemani, *Mendelian randomization: genetic anchors for causal inference in epidemiological studies*. Hum Mol Genet, 2014. **23**(R1): p. R89-98.
22. Burgess, S., et al., *Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors*. Eur J Epidemiol, 2015. **30**(7): p. 543-52.
23. Furberg, H., et al., *Genome-wide meta-analyses identify multiple loci associated with smoking behavior*. Nature Genetics, 2010. **42**(5): p. 441-U134.
24. Schumann, G., et al., *KLB is associated with alcohol drinking, and its gene product beta-Klotho is necessary for FGF21 regulation of alcohol preference*. 2016. **113**(50): p. 14372-14377.
25. Stringer, S., et al., *Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium*. Transl Psychiatry, 2016. **6**: p. e769.
26. Lawlor, D.A., et al., *Mendelian randomization: using genes as instruments for making causal inferences in epidemiology*. Stat Med, 2008. **27**(8): p. 1133-63.
27. Ehret, G.B., et al., *Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk*. Nature, 2011. **478**(7367): p. 103-9.
28. Bowden, J., et al., *Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator*. Genet Epidemiol, 2016. **40**(4): p. 304-14.
29. Bowden, J., G. Davey Smith, and S. Burgess, *Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression*. Int J Epidemiol, 2015. **44**(2): p. 512-25.
30. Burgess, S. and S.G. Thompson, *Interpreting findings from Mendelian randomization using the MR-Egger method*. Eur J Epidemiol, 2017. **32**(5): p. 377-389.
31. Burgess, S. and S.G. Thompson, *Avoiding bias from weak instruments in Mendelian randomization studies*. Int J Epidemiol, 2011. **40**(3): p. 755-64.
32. Bowden, J., et al., *Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic*. Int J Epidemiol, 2016. **45**(6): p. 1961-1974.
33. Bjorngaard, J., et al., *Cigarette smoking increases coffee consumption: findings from a Mendelian randomisation analysis*. International Journal of Epidemiology, in press.
34. Ware, J.J., et al., *Does coffee consumption impact on heaviness of smoking?* Addiction, 2017.
35. Irons, D.E., et al., *Mendelian randomization: A novel test of the gateway hypothesis and models of gene-environment interplay*. Development and psychopathology, 2007. **19**(4): p. 1181-1195.
36. Taylor, A.E., G. Davey Smith, and M.R. Munafo, *Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank*. Addiction, 2017.
37. Burgess, S., N.M. Davies, and S.G. Thompson, *Bias due to participant overlap in two-sample Mendelian randomization*. Genet Epidemiol, 2016. **40**(7): p. 597-608.

Table 1. Results of bi-directional, two-sample Mendelian randomization analyses testing for causal effects between use of nicotine, alcohol, caffeine and cannabis.

Exposure	Outcome	P-value threshold exposure	N SNPs	Inverse Variance Weighted (IVW) / Wald ratio			
				B	SE	OR	pval
Alcohol use	Caffeine use	5.00E-08	3	0.451	0.654		0.562
Alcohol use	Caffeine use	1.00E-05	40	0.059	0.127		0.647
Alcohol use	Cannabis use	5.00E-08	3	0.553	1.482	1.74	0.745
Alcohol use	Cannabis use	1.00E-05	62	0.455	0.260	1.58	0.085
Alcohol use	Smoking: Initiation	5.00E-08	3	-0.035	0.542	0.97	0.955
Alcohol use	Smoking: Initiation	1.00E-05	39	0.361	0.157	1.44	0.027
Alcohol use	Smoking: CPD	5.00E-08	3	-1.604	1.253		0.329
Alcohol use	Smoking: CPD	1.00E-05	39	-0.713	1.179		0.549
Caffeine use	Alcohol use	5.00E-08	4	0.029	0.019		0.235
Caffeine use	Alcohol use	1.00E-05	30	0.011	0.012		0.347
Caffeine use	Cannabis use	5.00E-08	4	-0.146	0.059	0.86	0.090
Caffeine use	Cannabis use	1.00E-05	28	-0.103	0.084	0.90	0.229
Caffeine use	Smoking: Initiation	5.00E-08	4	-0.029	0.048	0.97	0.590
Caffeine use	Smoking: Initiation	1.00E-05	31	-0.006	0.058	0.99	0.924
Caffeine use	Smoking: CPD	5.00E-08	4	-0.878	0.545		0.199
Caffeine use	Smoking: CPD	1.00E-05	31	0.256	0.382		0.508
Cannabis use	Alcohol use	1.00E-05	17	0.011	0.008		0.158
Cannabis use	Caffeine use	1.00E-05	11	-0.024	0.027		0.393
Cannabis use	Smoking: Initiation	1.00E-05	11	0.060	0.037	1.06	0.136
Cannabis use	Smoking: CPD	1.00E-05	11	-0.220	0.290		0.465
Smoking: Initiation	Alcohol use	1.00E-05	19	0.016	0.012		0.184
Smoking: Initiation	Caffeine use	1.00E-05	21	0.075	0.043		0.097
Smoking: Initiation	Cannabis use	1.00E-05	20	0.143	0.093	1.15	0.140
Smoking: CPD	Alcohol use	5.00E-08	1	-0.002	0.003		0.436
Smoking: CPD	Alcohol use	1.00E-05	11	0.000	0.002		0.755

Smoking: CPD	Caffeine use	5.00E-08	1	0.010	0.008		0.196
Smoking: CPD	Caffeine use	1.00E-05	11	0.015	0.005		0.012
Smoking: CPD	Cannabis use	5.00E-08	1	0.010	0.021	1.01	0.622
Smoking: CPD	Cannabis use	1.00E-05	11	0.011	0.008	1.01	0.225

P-value threshold exposure = p-value inclusion threshold of SNPs from genome-wide association studies (to

include in the genetic instrument); N SNPs = number of SNPs included in the genetic instrument; Inverse

Variance Weighted (IVW) / Wald ratio: columns represent results from the IVW analyses, unless the

genetic instrument consisted of 1 SNP in which case Wald ratio results are presented. B = risk coefficient

representing the change in outcome for a one-unit increase in the exposure variable; SE = standard error of

the B; OR = odds ratio representing the odds of the outcome variable with every unit increase of the exposure

variable.

Accepted Article