

Review

Mendelian Randomization Studies of Coffee and Caffeine Consumption

Marilyn C. Cornelis ^{1,*} and Marcus R. Munafò ²

¹ Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

² MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, UK Centre for Tobacco and Alcohol Studies, School of Psychological Science, University of Bristol, Bristol BS8 1TU, UK; marcus.munafò@bristol.ac.uk

* Correspondence: marilyn.cornelis@northwestern.edu; Tel.: +1-312-503-4548

Received: 31 August 2018; Accepted: 17 September 2018; Published: 20 September 2018



Abstract: Habitual coffee and caffeine consumption has been reported to be associated with numerous health outcomes. This perspective focuses on Mendelian Randomization (MR) approaches for determining whether such associations are causal. Genetic instruments for coffee and caffeine consumption are described, along with key concepts of MR and particular challenges when applying this approach to studies of coffee and caffeine. To date, at least fifteen MR studies have investigated the causal role of coffee or caffeine use on risk of type 2 diabetes, cardiovascular disease, Alzheimer’s disease, Parkinson’s disease, gout, osteoarthritis, cancers, sleep disturbances and other substance use. Most studies provide no consistent support for a causal role of coffee or caffeine on these health outcomes. Common study limitations include low statistical power, potential pleiotropy, and risk of collider bias. As a result, in many cases a causal role cannot confidently be ruled out. Conceptual challenges also arise from the different aspects of coffee and caffeine use captured by current genetic instruments. Nevertheless, with continued genome-wide searches for coffee and caffeine related loci along with advanced statistical methods and MR designs, MR promises to be a valuable approach to understanding the causal impact that coffee and caffeine have in human health.

Keywords: Mendelian Randomization; coffee; caffeine; behavior; causality; genetic epidemiology; epidemiological methods

1. Introduction

Coffee is one of the most widely consumed beverages in the world. Consumption patterns vary by country with larger per capita consumptions reported for Nordic countries, such as Finland (12.2 kg), Sweden (10.1 kg) and Norway (8.7 kg) compared to other countries such as Brazil (5.9 kg), Netherlands (5.3 kg), USA (4.5 kg), Australia (4.0 kg), Russia (1.7 kg), China (0.8 kg) and Turkey (0.7 kg) [1]. For most populations, regular coffee is the primary dietary source of caffeine; a psychostimulant also present in tea, cola, and cocoa products. Absorption and exposure to caffeine from these different sources is similar although a slight delay in absorption has been reported for cola and chocolate [2–4]. Roasted coffee also contains unique polyphenols (i.e., chlorogenic acid) and melanoidins that are major contributors to antioxidants in diet [5,6]. Boiled or unfiltered coffee contains diterpenoids, including cafestol and kahweol [7]. Trigonelline, magnesium, potassium, niacin, lignans, as well as heterocyclic amines and acrylamide have also been detected in the beverage [8–12]. With widespread popularity and availability of coffee, there is increasing public and scientific interest in the potential health consequences of its regular consumption. Traditional epidemiology has been fundamental to our increased knowledge on habitual coffee intake and health; but while a highly efficient and relevant

approach, it has several limitations that warrant consideration when interpreting the results [13]. Among these is establishing causal associations. The current perspective focuses on Mendelian Randomization (MR) approaches for determining a causal role of habitual coffee and caffeine intake on health. Because coffee and dietary caffeine intake are highly correlated we focus on both exposures. We first provide a brief review of coffee, caffeine and health. We follow with key concepts of the MR approach and particular challenges when applying it to studies of coffee and caffeine. Recent MR studies of coffee, caffeine and health are discussed, and we conclude with future directions for the field.

2. Coffee, Dietary Caffeine and Health

A recent umbrella review considered data from 201 meta-analysis of epidemiological studies of 67 unique health outcomes, and concluded that coffee likely has a beneficial role in reducing risk of type 2 diabetes (T2D), cardiovascular diseases (CVD), several cancers and Parkinson's disease (PD), but that high caffeine intake is likely harmful on pregnancy outcomes, such as low birth weight and pregnancy loss [14]. Overall, coffee consumption seems generally safe within usual levels of intake (i.e., at 3 to 4 cups a day) and more likely to benefit health than harm [14]. Rigorous reviews of caffeine toxicity conclude that consumption of up to 400 mg caffeine/day (equivalent to ~4 cups of coffee) in healthy adults, or 2.5 mg/kg/day for children and adolescents is not associated with overt adverse effects [15] and thus generally support the overall findings on habitual coffee intake and health [14]. Meanwhile, the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) lists caffeine intoxication and withdrawal as disorders, and have added 'caffeine use disorder' to 'Conditions for Further Study' [16]. Much of our knowledge pertaining to habitual coffee and caffeine intake on risk of chronic disease has been limited to observational research [14,15]. Inferring causality from observational data is difficult, due to potential residual confounding and reverse causality [17]. For example, in some populations coffee consumption is highly correlated with disease risk factors, such as smoking. Participants might acknowledge their true coffee behavior, but underreport their smoking behavior. As a consequence the coffee intake variable will continue to convey information about smoking even after adjustment for measures of smoking [18]. Coffee drinkers may also have reduced their coffee intake in light of disease symptoms or diagnosis, which might result in an apparent, but non-causal protective association between coffee and the disease [19]. Observational studies also provide no insight to mechanisms linking coffee to health. Coffee contains caffeine, but also hundreds of other chemicals that might benefit or harm health via different biological pathways [9]. Randomized trials of coffee consumption and disease outcomes would require long-term adherence to high or no coffee consumption, which is challenging given strong coffee consumption habits [20].

3. Mendelian Randomization (MR)

MR is a method of using the association of variation in genes with biomarkers or modifiable exposures to examine the causal effect of these biomarkers and exposures on disease outcomes in observational studies. The underlying principle of MR is that if a genetic variant alters the level of an exposure of interest, then this genetic variant should also be associated with disease risk and to the extent predicted by the effect of the genetic variant on the exposure [21,22]. According to Mendel's Law of Inheritance, alleles segregate randomly from parents to offspring. Thus, offspring genotypes are unlikely to be associated with confounders in the population. Moreover, germ-line genotypes are fixed at conception and so precede the observed variables, avoiding issues of reverse causation [23]. MR studies are often described as natural RCTs, but there are important differences [24]. For example, RCTs are usually of short duration while an individual's genetics generally reflect life-long exposures [21,24,25].

MR relies on a number of assumptions, in particular that the genetic variants(s): (1) Is associated with the modifiable exposure of interest, (2) is not associated with confounders of the exposure to outcome association and (3) only influences the outcome through the exposure of interest [17]. The first

assumption is the only one that can be formally tested, but MR methods and study designs have advanced much over the last few years and now include methods that are robust to potential violations of assumptions (2) and (3). It is increasingly widely used as a causal inference method in epidemiology. One-sample (genetics, exposure and outcome measured in the same sample) and two-sample (exposure and outcome measured in different samples) are the most common MR study designs. The latter is advantageous in situations where it is difficult to measure exposure and outcome in the same sample and can also be performed on publicly available genome-wide association study (GWAS) data (summary-level data). When possible, an instrument (genetic marker of exposure) that combines the effects of many SNPs is used to boost power while also addressing MR assumption violations (see below). The basic method for summary-level data, inverse-variance weighted (IVW), uses a fixed effects meta-analysis approach to combine the Wald ratio estimates of the causal effect (SNP-outcome effect divided by the SNP-exposure effect [26]) obtained from different SNPs, but assumes all SNPs are valid instruments or are invalid in such a way that the overall bias is zero [27,28]. The IVW is generally equivalent to the two-stage least squares estimate commonly used with individual level data.

4. Genetic Determinants of Coffee and Caffeine Consumption

Opportunities for MR studies of coffee and health have been made possible by the success of GWAS, which have identified multiple genetic variants associated with self-reported habitual coffee and caffeine consumption (Table 1) [29–33]. Loci near *ADORA2A*, *BDNF* and *SLC6A4* likely act directly on coffee drinking behavior by modulating the acute psychostimulant and rewarding properties of caffeine; driving factors for coffee drinking and caffeine use [34]. However, loci near *AHR*, *CYP1A2*, *POR*, and *ABCG2* generally present with the largest effect sizes and likely impact drinking behavior indirectly by altering the metabolism of caffeine and thus the physiological levels of this compound available for its psychostimulant effects. Only one locus is implicated in the sensory properties of coffee (*OR8U8*). Others have no obvious role in coffee or caffeine consumption, but have previously been associated with other traits in GWAS notably obesity, glucose and lipids [35–38]. GWAS and smaller follow-up studies have linked these loci to consumption of regular coffee, decaffeinated coffee, tea, total caffeine and water [31,39,40]. A subsequent GWAS of circulating caffeine metabolite levels further informed the roles of these loci in coffee and caffeine consumption behavior, but also identified variants near *CYP2A6* associated with paraxanthine-to-caffeine ratio (index for caffeine metabolism), that were nominally associated with drinking behavior [41]. Importantly, genetic variants leading to increased coffee/caffeine consumption associate with lower circulating caffeine levels and higher paraxanthine-to-caffeine ratio suggesting a fast caffeine metabolism phenotype. Thus, many of the loci affecting coffee and caffeine drinking behavior do so by modulating the physiological levels of caffeine.

5. Key Challenges to MR Studies of Coffee and Caffeine

Despite progress in the identification of robust genetic variants for coffee and caffeine consumption, efforts to apply these variants to MR studies of coffee and caffeine have been met with challenges, such as trait heterogeneity, pleiotropy and collider bias as discussed below. Limitations in the conduct and interpretation of MR studies more generally, along with potential solutions, have been reviewed in detail elsewhere [23,25,42], and include weak instrument bias, lack of reliable genetic instruments, population stratification, low statistical power (and therefore wide confidence intervals around causal estimates), linkage disequilibrium (LD) and the Winner's Curse phenomenon (i.e., the tendency for effect sizes in initial studies to be inflated).

Table 1. Genetic determinants of coffee and caffeine consumption [29–33].

Locus (Index SNP, Coffee/Caffeine Increasing Allele)	Closest Gene(s)	Encoded Protein(s): Function [UniProtKb]	Assoc. with Caffeine Metabolites *	Assoc. with Other Traits †	Hypothesized Link to Caffeine or Coffee Consumption
1q25.2 (rs574367, T)	<i>SEC16B</i>	SEC16 Homolog B, Endoplasmic Reticulum Export Factor: Required for secretory cargo traffic from the endoplasmic reticulum to the Golgi apparatus and for normal transitional endoplasmic reticulum organization.	$p > 0.05$	Y	None
2p25.3 (rs10865548, G)	<i>TMEM18</i>	Transmembrane Protein 18: Transcription repressor. Sequence-specific ssDNA and dsDNA binding protein, with preference for GCT end CTG repeats. Cell migration modulator, which enhances the glioma-specific migration ability of neural stem cells and neural precursor cells.	$p > 0.05$	Y	None
2p23.3 (rs1260326, C)	<i>GCKR</i>	Glucokinase regulatory protein (GKRP): Inhibits glucokinase by forming an inactive complex with this enzyme.	$\Downarrow\Downarrow$ $p < 1 \times 10^{-5}$	Y	Response to caffeine/coffee: May function in the glucose-sensing process of the brain that may influence central pathways responding to caffeine/coffee. Metabolism of caffeine: Inferred by association with caffeine metabolites
4q22 (rs1481012, A)	<i>ABCG2</i>	ATP-binding cassette sub-family G member 2: High-capacity urate exporter. Plays a role in porphyrin homeostasis and cellular export of heme and heme. May play an important role in the exclusion of xenobiotics from the brain. Implicated in the efflux of numerous drugs and xenobiotics.	\uparrow $p < 0.05$	Y	Metabolism of caffeine: Caffeine/metabolite efflux transporter.
7p21 (rs4410790 C, rs6968554, G)	<i>AHR</i>	Aryl hydrocarbon receptor: Ligand-activated transcriptional activator. Activates the expression of multiple phase I and II xenobiotic metabolizing enzymes. Involved in cell-cycle regulation and likely plays a role in the development/maturation of many tissues.	$\Downarrow\Downarrow$ $p < 5 \times 10^{-8}$	N	Metabolism of caffeine: Regulates <i>CYP1A2</i> expression.
7q11.23 (rs7800944, C)	<i>MLXIPL</i>	Carbohydrate-responsive element-binding protein: Transcriptional repressor.		Y	Response to caffeine/coffee: May regulate transcription of genes (e.g., <i>GCKR</i>) implicated in the response to caffeine.

Table 1. Cont.

Locus (Index SNP, Coffee/Caffeine Increasing Allele)	Closest Gene(s)	Encoded Protein(s): Function [UniProtKb]	Assoc. with Caffeine Metabolites *	Assoc. with Other Traits †	Hypothesized Link to Caffeine or Coffee Consumption
7q11.23 (rs17685, A)	<i>POR</i>	NADPH-cytochrome P450 reductase: Required for electron transfer from NADP to cytochrome P450 in microsomes and can also facilitate electron transfer to heme oxygenase and cytochrome B5.	↓ $p < 0.05$	N	Metabolism of caffeine: Required for CYP1A2 catalytic activity.
11p13 (rs6265, C)	<i>BDNF</i>	Brain-derived neurotrophin factor: During development, promotes survival and differentiation of selected neuronal populations of the PNS and CNS. Major regulator of synaptic transmission and plasticity at adult synapses in many regions of the CNS.	$p > 0.05$	Y	Response to caffeine: Modulates neurotransmitters potentially mediating the rewarding response to caffeine.
11q12.1 (rs597045, A)	<i>OR8U8</i>	Olfactory Receptor Family 8 Subfamily U Member 8: Odorant receptor	$p > 0.05$	N	Smell/taste perception of coffee
14q12 (rs1956218, G)	<i>AKAP6</i>	A-Kinase Anchoring Protein 6: Binds to type II regulatory subunits of protein kinase A and anchors/targets them to the nuclear membrane or sarcoplasmic reticulum. May act as an adapter for assembling multiprotein complexes.	$p > 0.05$	N	None
15q24 (rs2470893 T, rs2472297, T)	<i>CYP1A1, CYP1A2</i>	Cytochrome P450 1A1/2: Cytochromes P450 are a group of enzymes involved in NADPH-dependent electron transport pathways. They oxidize a variety of compounds, including steroids, fatty acids, and xenobiotics.	↓↓ $p < 5 \times 10^{-8}$	N	Metabolism of caffeine: CYP1A2 metabolizes >95% of caffeine.
17q11.2 (rs9902453, G)	<i>EFCAB5</i> <i>SLC6A4</i>	EF-hand calcium-binding domain-containing protein 5: Unknown Sodium-dependent serotonin transporter: In CNS, regulates serotonergic signaling via transport of serotonin molecules from the synaptic cleft back into the presynaptic terminal for reuse.	$p > 0.05$	N	Response to caffeine/coffee: Serotonin may mediate the rewarding response to caffeine.
18q21.32 (rs66723169, A)	<i>MC4R</i>	Melanocortin 4 Receptor: Receptor specific to the heptapeptide core common to adrenocorticotrophic hormone and alpha-, beta-, and gamma-MSH. Plays a central role in energy homeostasis and somatic growth.	$p > 0.05$	Y	None
22q11.23 (rs2330783, G)	<i>SPECCIL-ADORA2A</i>	Adenosine A2a Receptor: Receptor for adenosine. The activity of this receptor is mediated by G proteins, which activate adenylyl cyclase.	↑ $p < 0.05$	N	Response to caffeine/coffee: Caffeine blocks this receptor, which mediates some of the psychostimulant effects of caffeine.

* SNP is associated with (i) higher blood levels of caffeine (↑); (ii) lower blood levels of caffeine (↓); or (iii) lower blood levels of caffeine and higher paraxanthine-to-caffeine ratio (↓↓).

† GWAS (genome-wide association study) catalogue traits unrelated to caffeine or coffee. Y, Yes; N, No.

5.1. Trait Heterogeneity

The most comprehensive (and therefore powerful) genetic instrument employed in an MR study of coffee will reflect multiple aspects of coffee drinking behavior (Table 2), such as caffeine metabolism, reward-response and potentially taste. Such heterogeneity does not preclude causal inference, but it does limit the ability to infer causality for particular dimensions of coffee (e.g., caffeine vs non-caffeine) and makes interpretation of MR analyses more difficult [23,25]. An instrumental variable (IV) that narrows in on a particular aspect of coffee drinking might also face issues of interpretation. For example, genetically-inferred ‘fast’ and ‘slow’ caffeine metabolizers may consume different amounts of the same type of coffee, but their circulating caffeine levels may not be different. However, circulating levels of non-caffeine constituents of coffee *will* differ. Alternatively, given the same amount and type of coffee consumed, slow caffeine metabolizers will, on average, have higher circulating caffeine levels than fast caffeine metabolizers. Circulating levels of non-caffeine constituents will generally be the same. Because most of the SNPs associate with caffeine intake, and not exclusively coffee intake, the genetic instrument for coffee might also reflect exposure to other dietary sources of caffeine, which might confound or mask any causal relationship between coffee and outcome [43]. Although MR studies are thought to be relatively protected against exposure measurement error, this is less likely to be the case for an MR study of coffee or caffeine [20]. For example, the genetic predisposition to drink coffee, due to an increased caffeine metabolism might also impact preference for regular strong coffee over other coffee types. Taken together, it is important to specify the hypothesis being tested a priori, select the optimal IV and sample for analysis, and consider alternate explanations for positive or null results.

Table 2. Mendelian Randomization (MR) studies of coffee and caffeine consumption.

Study	Outcome	Instrumental Variable (IV)	Design & Approach	Results	Interpretation	Limitations Reported
Nordestgaard et al. 2015 [44]	Obesity, metabolic syndrome, T2D and related measures (BMI, WC, height, weight, SBP, DBP, TGs, TC, HDL, glucose)	5-SNPs <i>AHR, CYP1A2</i> Score and single SNPs	One-sample Individual-level data 2SLS $n \leq 93,179$ Copenhagen General Population Study (CGPS) and the Copenhagen City Heart Study (CCHS). Summary-level data Wald ratio, IVW T2D only DIAGRAM ($n \leq 78,021$)	Observational: Coffee significantly reduced risk of obesity, metabolic syndrome and T2D Coffee significantly increased BMI, WC, weight, height, SBP, DBP, TGs, and TC and decreased HDL SNP-outcome: NS Similar results when individuals were stratified into coffee drinkers and coffee abstainers however, among those without coffee intake, blood pressure was lower with higher coffee-intake allele score	No evidence supporting a causal relationship between coffee and outcomes	Underpowered IV Pleiotropy Collider Bias
Nordestgaard & Nordestgaard, 2016 [43]	CVD (IHD, IS, IVD) All-cause and CVD mortality	2-SNPs <i>AHR, CYP1A2</i> Score and single SNPs	One-sample Individual-level data 2SLS $n \leq 112,509$ CGPS, CCHS and Copenhagen Ischaemic Heart Disease Study (CIHDS) 3822 IHD cases 1708 IS cases 4971 IVD cases 971 CVD deaths 5422 total deaths Summary-level data Wald ratio, IVW IHD only Cardiogram ($n = 80,517$) and C4D ($n = 30,433$)	Observational: U-shaped association between coffee intake and IHD, IS, IVD and all-cause mortality. Lowest risk with medium coffee intake compared with no coffee intake. SNP-outcome: NS Similar results when individuals were stratified into coffee abstainers, coffee drinkers, coffee drinkers excluding tea and cola drinkers.	No evidence supporting a causal relationship between coffee and outcomes	Underpowered IV Pleiotropy Collider Bias (stratified analysis) Confounding by other caffeine containing-beverages Cannot rule out non-linear effects of coffee on outcomes
Kwok et al., 2016 [45]	T2D, IHD, depression, Alzheimer's disease, lipids, glycemic traits, adiposity or adiponectin	9-SNPs <i>AHR, CYP1A2(2), GCKR, MLXIPL, POR, EFCAB5, BDNF, ABCG2</i> 5 SNPs <i>AHR, CYP1A2(2), POR, EFCAB5</i> 3 SNPs <i>AHR, CYP1A2(2)</i>	Two-sample Summary-level data Multiple published GWAS WME	9 SNPs: \uparrow T2D, \downarrow TGs, \uparrow BMI, \uparrow WHR, \uparrow IR 5 SNPs: NS 3 SNPs: NS	No evidence supporting a causal relationship between coffee and outcomes	Confounding (Population stratification) Pleiotropy Cannot rule out non-linear effects of coffee on outcomes

Table 2. Cont.

Study	Outcome	Instrumental Variable (IV)	Design & Approach	Results	Interpretation	Limitations Reported
Treur et al., 2016 [46]	Smoking behavior Coffee intake Caffeine use	1-SNP for smoking heaviness (<i>CHRNA3</i>) 8-SNP score for coffee intake <i>AHR, CYP1A2, GCKR, MLXIPL, POR, EFCAB5, BDNF, ABCG2</i>	Individual-level data Bivariate genetic modelling (SEM) $n = 10,368$ current smoking (y/n) caffeine use (high/low) coffee use (high/low) Bidirectional MR Regression analyses $n = 12,319$ Self-reported caffeine use (mg/day), coffee use (cups/day), cigs/day, smoking initiation and persistence Summary-level data LD score regression CCGC Tobacco, Alcohol and Genetics Consortium (TAG): cigs/day, smoking initiation and persistence $n \leq 38,181$	Bivariate genetic modelling Current smoking-coffee intake: $G r = 0.47, E r = 0.30$ Current smoking-caffeine use: $G r = 0.44, E r = 0.00$ MR: NS LD score regression Smoking heaviness- coffee intake: $r = 0.44$ Smoking initiation-coffee intake: $r = 0.28$ Smoking persistence-coffee intake: $r = 0.25$	Genetic factors explain most of the association between smoking and caffeine consumption. Quitting smoking may be more difficult for heavy caffeine consumers, given their genetic susceptibility.	Underpowered Pleiotropy
Taylor et al., 2017 [47]	Prostate cancer (PC) risk and progression	2-SNPs <i>AHR, CYP1A2</i>	Individual-level data Two-sample MR Regression analyses + meta-analysis Practical consortium ($n = 46,687$) 4 studies GS-coffee GS-tea GS-(tea + coffee) 23 studies GS-PC GS-PC stage GS-PC grade GS-mortality	Significant GS-coffee, GS-tea and GS-(tea + coffee) GS-PC grade ($p = 0.02$)	No clear evidence supporting a causal relationship between coffee and outcomes	Between-study heterogeneity in case definition Imprecise IV Pleiotropy Underpowered

Table 2. Cont.

Study	Outcome	Instrumental Variable (IV)	Design & Approach	Results	Interpretation	Limitations Reported
Ware et al., 2017 [48]	Smoking heaviness, cigs/day	8-SNP GS <i>AHR, CYP1A2, GCKR, MLXIPL, POR, EFCAB5, BDNF, ABCG2</i> 6-SNP GS <i>AHR, CYP1A2, GCKR, MLXIPL, POR, EFCAB5</i> 2-SNP GS <i>AHR, CYP1A2</i>	2-sample MR Summary-level data IVW, WME CCGC TAG GWAS Cotinine levels ($n = 4548$) [in vitro experiments] Individual-level data (replication, $n = 8072$ smokers who drink coffee) IVW, WME	Each cup of coffee/day lead to a decrease in 1.5 (8 SNPs), 1.7 (6 SNPs) or 2.0 (2 SNPs) cigs/day. Coffee did not influence cotinine levels. Coffee did not influence cigs/day in replication sample.	Coffee intake is unlikely to have a major causal impact on cigarette smoking	Pleiotropy Underpowered replication Underpowered IV
Bjorngaard et al., 2017 [49]	Coffee intake (cups/day, sensitivity analysis: Any vs. none) Tea intake (cups/day, sensitivity analysis: Any vs. none) Smoking status (never, former, current) Smoking heaviness (cigs/day)	1-SNP (<i>CHRNA3</i>) for smoking heaviness 2-SNPs (<i>AHR, CYP1A2</i>) for coffee intake GS	Individual-level data Bidirectional MR Regression analyses + meta-analysis UK biobank ($n \leq 114,029$) HUNT ($n \leq 56,664$) CGPS ($n \leq 78,650$) coffee or tea drinkers only	Observational Former & current smoking associated with higher coffee consumption (not tea) vs. never smokers. Among smokers: Each cig/day increased coffee and tea intake; stronger for coffee MR SMK-SNP associated with coffee intake in current or ever smokers only Coffee-SNP not associated with smoking behavior	Higher cigarette consumption causally increases coffee intake.	Underpowered to rule out causal coffee → smoking association. UK Biobank non-representative sample Collider bias: (i) if selection into the sample is related to both coffee and smoking (ii) via smoking stratification Phenotype measurement error
Larsson et al., 2017 [50]	Alzheimer's Disease (AD)	5-SNP GS <i>AHR, CYP1A2, MLXIPL, POR, EFCAB5</i> (coffee and 23 other exposures tested)	Summary-level data 2-sample MR IVW, WME, MR Egger CCGC International Genomics of Alzheimer's Project ($n = 17,009$ cases, 37,154 controls)	Suggestive association between coffee GS and increased risk of AD ($p = 0.01$)	Suggestive causal relationship between coffee and AD risk, but in opposite direction to that expected based on observational studies.	None.
Verweij et al., 2018 [51]	Causal associations between nicotine, alcohol, caffeine, and cannabis use	Polygenic scores ($p < 5 \times 10^{-8}$ or $p < 1 \times 10^{-5}$) for each exposure	Summary-level data two-sample bidirectional MR IVW, Wald ratio Multiple published GWAS	Smoking cigs/day—caffeine use ($p = 0.01$) Alcohol use: Smoking initiation ($p = 0.03$)	Little evidence for causal relationships between nicotine, alcohol, caffeine, and cannabis use, but may suggest a common liability model (shared genetics)	Imprecise IV GWAS sample overlap (bias to null)

Table 2. Cont.

Study	Outcome	Instrumental Variable (IV)	Design & Approach	Results	Interpretation	Limitations Reported
Ong et al., 2017 [52]	Epithelial ovarian cancer	4-SNP GS (coffee IV) <i>ABCG2, AHR, CYP1A2, POR</i> 2-SNP GS (caffeine IV) <i>AHR, CYP1A2</i>	Summary-level data Two-sample MR Wald-type ratio estimator CCGC Ovarian Cancer Association Consortium (<i>n</i> = 44,062, 20,683 cases)	NS	No evidence supporting a causal relationship between coffee/caffeine and outcome	MR Assumption 3 not confirmed Not generalizable to non-European populations. Underpowered or imprecise IV Cannot rule out non-linear effects of coffee/caffeine on cancer
Larsson et al., 2018 [53]	Gout	5-SNPs <i>AHR, CYP1A2, MLXIPL, POR, EFCAB5</i>	Summary-level data 2-sample MR IVW, WME, MR Egger CCGS Serum Uric acid GWAS (<i>n</i> = 110,347) Gout GWAS (2115 cases and 67,259 controls).	<i>CYP1A2</i> and <i>MLXIPL</i> SNPs inversely associated with uric acid Combined MR: significant inverse relationship ($p = 7.9 \times 10^{-6}$) All but <i>AHR</i> SNP associated with lower gout risk. Combined MR: significant inverse relationship ($p = 0.005$)	Supports causal inverse association between coffee intake and risk of gout.	None
Treur et al., 2018 [54]	Sleep behaviors (sleep duration, chronotype and insomnia complaints)	IV threshold $p < 5 \times 10^{-8}$ 4 SNPs (<i>POR, AHR, CYP1A2, MXLIPL</i>) $p < 5 \times 10^{-5}$ 4 SNPs plus 23 SNPs	Summary-level data Two-sample bidirectional MR IVW, LD score regression CCGC Caffeine metabolite GWAS Sleep GWAS	MR: NS LD score regression: NS	No evidence for causal relationship between habitual coffee intake and sleep behaviors.	Underpowered LD score regression using caffeine metabolite GWAS Phenotype measurement error
Noyce et al., 2018 [55]	Parkinson's Disease (PD)	Morning person primary exposure (15 SNPs) coffee secondary exposure (4-SNPs, <i>AHR, BDNF, POR, CYP1A2</i>)	Summary-level data Two-sample MR IVW CCGC Morning person GWAS (<i>n</i> = 89,283) PD GWAS (13,708 cases, 95,282 controls)	Morning person MR: $p = 0.01$ Coffee MR: NS	Along with published RCT results, findings suggest that caffeine may neither prevent PD occurring nor be of benefit in those with the condition.	Use of summary-level data does not allow adjustment for potential confounding factors.

Table 2. Cont.

Study	Outcome	Instrumental Variable (IV)	Design & Approach	Results	Interpretation	Limitations Reported
Zhou et al. 2018 [56]	Cognitive function composite global cognition and memory scores	2-SNPs <i>AHR, CYP1A2</i> Other SNPs (secondary analysis)	Individual-level data <i>n</i> = 415,530 (300,760 coffee drinkers) from 10 meta-analyzed European ancestry cohorts. Genetic analysis performed under different levels of habitual coffee intake (1–4 and ≥4 cups/day. Negative control: Non-coffee drinkers.	Observational: No overall association between coffee intake and global cognition and memory. SNP-outcome: NS	Study provides no evidence to support beneficial or adverse long-term effects of coffee intake on global cognition or memory.	Pleiotropy. Caution when interpreting coffee IV
Lee, 2018 [57]	Osteoarthritis	4 SNPs, <i>POR, CYP1A2, NRCAM, NCALD</i>	Summary-level data Two-sample MR IVW, WME, MR-Egger regression CCGC + Amin et al. 2012 (<i>n</i> = 18,176) Osteoarthritis GWAS (7410 cases, 11,009 controls)	IVW: <i>p</i> = 0.03 WME: <i>p</i> = 0.05 MR Egger: NS (however, no pleiotropy was evident)	Results suggest that coffee consumption is causally associated with an increased risk of osteoarthritis.	Underpowered or imprecise IV Results limited to populations of European ancestry and limited to osteoarthritis in the knee and hip

AD—Alzheimer’s disease; BMI—body mass index; CCGC—Coffee and Caffeine Genetics Consortium; DBP—diastolic blood pressure; DIAGRAM—Diabetes Genetics Replication and Meta-analysis; GS—genetic (SNP) score; HDL—high-density lipoprotein; IHD—ischaeic heard disease, IS—ischaeic stroke, IVD—ischaeic vascular disease, IVW—inverse-variance weighted meta-analysis, NS—non-significant; PC—prostate cancer; PD—Parkinson’s Disease; SBP—systolic blood pressure; T2D—type 2 diabetes; TC—total cholesterol; TGs—triglycerides; WC—waist circumference; WME—weighted median estimate.

5.2. Pleiotropy

Pleiotropy can violate MR assumption 3, which requires that the genetic variant only influences the outcome through the exposure of interest. Vertical pleiotropy does not violate MR assumption 3 and occurs when the genetic variant is associated with a factor on the pathway between the exposure and outcome, but only because of its effect on the exposure [58]. Horizontal (or biological) pleiotropy occurs when a genetic variant is associated with multiple exposures or traits and is therefore a violation of MR assumption 3 [17,58]. Seven of the fourteen loci associated with coffee or caffeine consumption are also associated with other traits based on GWAS [35] (Table 1). Whether this results from horizontal pleiotropy or a true causal relationship between coffee and these other traits is unclear. Nevertheless, since it is not possible to prove assumption 3 holds for all SNPs in an MR study its becoming common practice to implement extensions of the basic MR methodology that detect the presence of pleiotropy and account for it in causal estimates of the exposure [59]. Random effects IVW or weighted generalized linear regressions are simple options [22,60,61], but common methods that explicitly account for pleiotropy include MR-Egger regression [62], and the weighted-median estimate [63]. Newer methods include MR-PRESSO [64] and generalized summary MR (GSMR) [65]. Each approach relies on different (and largely uncorrelated) assumptions, and therefore the use of multiple approaches allows triangulation; if all provide consistent causal estimates we can be more confident that a true causal effect exists.

5.3. Collider Bias

When individual-level data are available, a common strategy is to restrict SNP-outcome analysis to coffee drinkers arguing that the SNPs are associated with coffee drinking (heaviness) and thus causal relationships should only be observed among coffee drinkers (a form of gene-environment interaction) [43,44,48,49,56,59,66]. SNP-outcome associations among non-drinkers ('negative control sample') would suggest a violation in at least one of the assumptions [59,66]. However, this strategy introduces potential for collider bias given that several loci associated with coffee intake also distinguish between non-drinker and heavy coffee drinkers [31]. Collider bias occurs when the exposure and outcome of interest independently influence a third risk factor, and this third risk factor is conditioned upon, either through statistical adjustment or stratification [67–69]. This bias will also apply to the genetic correlates of the exposure and outcome. Indeed, MR studies of coffee intake among the Copenhagen population provided evidence for collider bias [43,44]. For example, among coffee-abstainers, the genetic IV for coffee intake was inversely associated with age. Since age was a risk factor for the outcome and was strongly associated with coffee intake, but among coffee consumers only, the IV-age association in the 'negative control sample' likely arises from collider bias [43].

6. MR Studies of Coffee, Caffeine and Health

Table 2 summarizes all MR studies of coffee or caffeine and health outcomes published to-date. Studies are in descending order by date of publication (column 1). For each study we extracted the outcome of interest (column 2), the genetic variants used as the IV (column 3), the basic design and approach (column 4), main results (column 5), interpretation or overarching conclusion of the study (column 6) and limitations as acknowledged by study authors (column 7). With one exception [57], all study IVs included at least SNPs near *CYP1A2* and *AHR*—the strongest and most robust variants linked to coffee drinking behavior and caffeine metabolite levels (Table 1). Primary analysis was conducted using predominately regression analyses or IVW meta-analysis for multi-SNP analysis. These were generally followed by weighted median estimates and MR-Egger regressions to address potential assumption violations. In most studies, the exposure of interest was simply defined as coffee consumption or caffeine use. Data from the GWAS of coffee consumption among 91,462 coffee drinkers in the Coffee and Caffeine Genetics Consortium (CCGC) [31] were used in all summary-level data analysis.

Epidemiological studies report a consistent inverse linear association between coffee consumption and T2D [14], which extends to decaffeinated coffee. This is typically interpreted as evidence for non-caffeine constituents of coffee underlying the coffee-T2D relationship [14]. Two studies, using individual-level and summary-level data for up to ~170,000 participants (26,000 T2D cases) provided no evidence in support of a causal association between coffee intake and T2D risk [44,45], which also extended to measures of adiposity, blood pressure, lipid and glucose metabolism [44,45]. Nordestgaard and colleagues [44] additionally examined a BMI IV (SNPs in/near *FTO*, *MC4R* and *TMEM18*) to examine potential reverse causation from BMI to coffee intake, and as a positive control for risk of T2D. The coffee-intake IV was not linked to BMI, but the BMI-IV was positively associated with coffee intake. Interestingly, SNPs included in the BMI-IV were recently shown to associate with coffee consumption in GWAS (Table 1) [33] and so possibly relate to reward mechanisms (the causal pathway) relevant to coffee drinking behavior and obesity and not adiposity per se [33].

Epidemiological studies also suggest coffee intake may reduce risk of CVD, CVD-mortality and all-cause mortality, but with greatest risk reduction with 3 to 5 cups/day (i.e., a non-linear association) [14]. Nordestgaard and Nordestgaard [43] examined all three of these outcomes in 112,509 Danes and observed a similar pattern of benefits associated with coffee consumption over a 6 year follow-up, but no evidence for causality. In the subgroup of coffee drinkers they noted strong positive and plausible LDL-SNP and HRT-SNP associations, but could not rule-out that such associations could have resulted from collider bias [43].

Caffeine, nicotine, alcohol, and cannabis use are highly correlated behaviors [70]. Potential mechanisms include shared genetic and/or shared environmental factors (i.e., common liability) or a causal influence of one on the other [71]. The co-occurrence of coffee/caffeine use with other substance use behaviors has been investigated in four MR studies [46,48,49,51]. Three of these studies employed bidirectional MR [46,49,51], in which IVs for each substance use were used to evaluate causal effects and their direction [23,72]. The first study focused on the association between smoking and caffeine using three approaches: Bivariate genetic modelling in a twin sample, LD score regression with summary level-data and bidirectional MR analysis using individual-levels data [46]. The results suggested shared genetic factors for caffeine/coffee intake and smoking behavior, rather than a causal influence of one behavior on the other. Ware and colleagues [48] specifically focused on the causal role of coffee consumption on smoking heaviness. Two-sample MR analyses indicated that heavier coffee consumption might lead to *reduced* heaviness of smoking. However, their in vitro experiments, and attempt to replicate in the UK Biobank sample of smokers who drank coffee, did not support these initial causal findings, and overall were not consistent with the direction of association reported in observational analysis. Bjorngaard and colleagues [49] also examined coffee and tea drinkers from three population studies using bidirectional MR and provided evidence for a causal relationship of smoking heaviness on coffee and tea intake, but not vice versa. Finally, Verweij and colleagues [51] examined causal relationships among caffeine, smoking, as well as alcohol, and cannabis use with a variation of bidirectional MR that used 'polygenic scores'. The latter relaxes the significance threshold for GWAS to produce a stronger instrument, but also runs the risk of vertical pleiotropy [59]. Their findings did not support the hypothesis that causal relationships explain the co-occurrence of use of different substances, but are consistent with a common liability model [51].

Alzheimer's Disease (AD) was investigated by Kwok and colleagues [45], and Larsson and colleagues [50], using the same summary-level data, but employed different multi-SNP IVs. Larsson and colleagues [50] used an IV with SNPs for *AHR*, *CYP1A2*, *MLXIPL*, *POR* and *EFCAB* and reported a suggestive causal relationship between coffee and AD risk, but in the opposite direction to that expected based on observational data. Kwok and colleagues [45], whom did not include the *MLXIPL* SNP in their IV, reported no evidence for a causal relationship. A causal relationship between coffee and cognitive function was also not supported by a separate MR [56]. The latter accounted for the potential non-linear association between coffee and cognitive function by conducting analysis by different levels of coffee intake. An association among non-coffee consumers served as a negative

control sample. While collider bias was not acknowledged as a limitation, they noted caution when interpreting their results as the instruments indexing greater caffeine consumption may reflect a faster rate of caffeine clearance, and hence a lower (rather than higher) circulating level of bioactive caffeine [56].

Although data are limited, coffee intake has been linked to lower risk of gout [14]. Larsson and colleagues [53] examined the causal association between coffee and gout, as well as uric acid, a related biomarker. The five SNP-IV (excluding the *ABCG2* SNP, which associates with uric acid) was inversely related to both gout risk and uric acid levels, supporting a causal relationship between coffee drinking and gout.

MR studies have failed to support a causal association between coffee/caffeine intake and epithelial ovarian cancer [52], prostate cancer [47], sleep behaviors [54] and Parkinson's disease (PD) [55]. The latter finding is in marked contrast to consistent observational and animal experimental data suggesting coffee and caffeine are protective for PD, but rather align with RCTs and suggest "caffeine may neither prevent PD occurring nor be of benefit in those with the condition" [55]. The authors nevertheless noted that potentially causal effects of coffee may not occur exclusively through caffeine [55], suggesting their IV aimed to capture caffeine exposure rather than coffee drinking per se. The most recent coffee MR was applied to osteoarthritis [57] and supported a causal positive relationship between coffee and this outcome. However, the selection of SNPs for the study was unclear and no human observational study has examined coffee and osteoarthritis, so that the findings are largely hypothesis-generating.

Taken together, at least fifteen studies to date have investigated the causal role of coffee or caffeine use in T2D, CVD, AD and cognition, PD, gout, osteoarthritis, cancers, sleep and other substance use behaviors. Single studies investigated and provided support for a causal role of coffee in reducing risk of gout [14] and increasing risk of osteoarthritis [57]. Four studies examined the co-occurrence of caffeine use and other substances with conflicting results [46,48,49,51]. For the remaining outcomes, studies did not provide clear support for a causal role of coffee or caffeine, but often acknowledged limitations (such as low statistical power, pleiotropy and collider bias), such that a causal role cannot yet be ruled out.

7. Future Directions

There is continued enthusiasm for understanding the causal role of coffee and caffeine in health. Thus far, most outcomes of interest have been investigated by single studies and thus the significant and null findings warrant confirmation in independent studies. Many outcomes, for which coffee and caffeine have been implicated, have yet to be investigated [14]. Methodological challenges, such as insufficient power, pleiotropy and collider bias are commonly acknowledged. However, conceptual challenges arising from the different aspects of coffee/caffeine use captured by genetic instruments warrant careful consideration going forward. With continued investment in GWAS it may be possible to parse variants related to non-caffeine aspects of coffee from those related to caffeine providing opportunities to identify the causal elements of coffee per se, rather than coffee drinking behavior. The increasing availability of large individual-level data sets and advanced statistical methods means that more sophisticated MR designs might also be considered. For example, the use of polygenic scores might be optimized using the MR robust adjusted profile score (MR-RAPS) method, which weights each variant differently based on effect size and precision of the SNP-exposure association [62]. Given the co-occurrence of coffee drinking and smoking, a factorial MR may be an attractive approach to study the combined causal effects (i.e., interaction) of these behaviors on disease [22]. Individuals can be allocated into either a high or low-SNP score for coffee and then each group further allocated into either a high or low-SNP score for smoking. The causal estimates for each of the resulting four groups on disease could then be determined. A two-step MR may also be used to assess whether an intermediate trait, say a biomarker or metabolite, acts as a causal mediator between coffee drinking and an outcome [73,74]. An IV for coffee drinking is first used to estimate the causal effect of coffee drinking

on the potential mediator (step 1). IVs for the potential mediator are then used to assess the causal effect of the mediator on the outcome (step 2). Evidence of association in both steps implies some degree of mediation of the association between coffee drinking and the outcome by the intermediate variable. Finally, multivariable MRs allow multiple exposures to be examined simultaneously, and provide an effect estimate of one conditional on the other (e.g., effects of coffee consumption conditional on circulating caffeine levels) [75]. These alternate MR designs will still require careful attention to challenges and limitations discussed above.

Multiple statistical methods to accommodate different MR violations combined with replication studies and other mechanistic studies will be necessary to support stronger causal relationship between coffee or caffeine intake and health [59]. GWAS of more refined coffee drinking behaviors, and circulating metabolite markers of coffee intake will also be important, but the collection of such data on a large scale will be needed first. Nevertheless, in light of the rapid pace, in which advancements are being made in these areas, MR promises to be an increasingly valuable approach to understanding the causal impact that coffee and caffeine have in human health.

Author Contributions: M.C.C. and M.R.M. conceptualized the paper. M.C.C. wrote the first draft of the paper. All authors revised and approved the final the manuscript.

Acknowledgments: This work was funded by the National Institute on Aging (K01AG053477 to M.C.C.) and a Benjamin Meaker Visiting Professorship (to M.C.C.).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. International Coffee Organization. Trade Statistics. Available online: http://www.ico.org/profiles_e.asp (accessed on 1 August 2018).
2. Marks, V.; Kelly, J. Absorption of caffeine from tea, coffee, and coca cola. *Lancet* **1973**, *301*, 827. [[CrossRef](#)]
3. Nehlig, A. Interindividual differences in caffeine metabolism and factors driving caffeine consumption. *Pharmacol. Rev.* **2018**, *70*, 384–411. [[CrossRef](#)] [[PubMed](#)]
4. White, J.R., Jr.; Padowski, J.M.; Zhong, Y.; Chen, G.; Luo, S.; Lazarus, P.; Layton, M.E.; McPherson, S. Pharmacokinetic analysis and comparison of caffeine administered rapidly or slowly in coffee chilled or hot versus chilled energy drink in healthy young adults. *Clin. Toxicol.* **2016**, *54*, 308–312. [[CrossRef](#)] [[PubMed](#)]
5. Scalbert, A.; Williamson, G. Dietary intake and bioavailability of polyphenols. *J. Nutr.* **2000**, *130*, 2073S–2085S. [[CrossRef](#)] [[PubMed](#)]
6. Yanagimoto, K.; Ochi, H.; Lee, K.G.; Shibamoto, T. Antioxidative activities of fractions obtained from brewed coffee. *J. Agric. Food Chem.* **2004**, *52*, 592–596. [[CrossRef](#)] [[PubMed](#)]
7. Urgert, R. Levels of the cholesterol-elevating diterpenes cafestol and kahweol in various coffee brews. *J. Agric. Food Chem.* **1995**, *43*, 2167–2172. [[CrossRef](#)]
8. Milder, I.E.; Arts, I.C.; van de Putte, B.; Venema, D.P.; Hollman, P.C. Lignan contents of dutch plant foods: A database including lariciresinol, pinoresinol, secoisolariciresinol and matairesinol. *Br. J. Nutr.* **2005**, *93*, 393–402. [[CrossRef](#)] [[PubMed](#)]
9. Spiller, M.A. The chemical components of coffee. In *Caffeine*; Spiller, G.A., Ed.; CRC: Boca Raton, FL, USA, 1998; pp. 97–161.
10. Andrzejewski, D.; Roach, J.A.; Gay, M.L.; Musser, S.M. Analysis of coffee for the presence of acrylamide by lc-ms/ms. *J. Agric. Food Chem.* **2004**, *52*, 1996–2002. [[CrossRef](#)] [[PubMed](#)]
11. Minamisawa, M.; Yoshida, S.; Takai, N. Determination of biologically active substances in roasted coffees using a diode-array hplc system. *Anal. Sci.* **2004**, *20*, 325–328. [[CrossRef](#)] [[PubMed](#)]
12. Ludwig, I.A.; Mena, P.; Calani, L.; Cid, C.; Del Rio, D.; Lean, M.E.; Crozier, A. Variations in caffeine and chlorogenic acid contents of coffees: What are we drinking? *Food Funct.* **2014**, *5*, 1718–1726. [[CrossRef](#)] [[PubMed](#)]
13. Cornelis, M.C. Toward systems epidemiology of coffee and health. *Curr. Opin. Lipidol.* **2015**, *26*, 20–29. [[CrossRef](#)] [[PubMed](#)]
14. Poole, R.; Kennedy, O.J.; Roderick, P.; Fallowfield, J.A.; Hayes, P.C.; Parkes, J. Coffee consumption and health: Umbrella review of meta-analyses of multiple health outcomes. *BMJ* **2017**, *359*, j5024. [[CrossRef](#)] [[PubMed](#)]

15. Wikoff, D.; Welsh, B.T.; Henderson, R.; Brorby, G.P.; Britt, J.; Myers, E.; Goldberger, J.; Lieberman, H.R.; O'Brien, C.; Peck, J. Systematic review of the potential adverse effects of caffeine consumption in healthy adults, pregnant women, adolescents, and children. *Food Chem. Toxicol.* **2017**, *109*, 585–648. [[CrossRef](#)] [[PubMed](#)]
16. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed.; American Psychiatric Publishing: Arlington, VA, USA, 2013.
17. Davey Smith, G.; Hemani, G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **2014**, *23*, 89–98. [[CrossRef](#)] [[PubMed](#)]
18. Leviton, A. Coffee consumption and residual confounding. *Epidemiology* **1996**, *7*, 110. [[CrossRef](#)] [[PubMed](#)]
19. Soroko, S.; Chang, J.; Barrett-Connor, E. Reasons for changing caffeinated coffee consumption: The rancho bernardo study. *J. Am. Coll. Nutr.* **1996**, *15*, 97–101. [[CrossRef](#)] [[PubMed](#)]
20. van Dam, R. Can 'omics' studies provide evidence for causal effects of coffee consumption on risk of type 2 diabetes? *J. Int. Med.* **2018**, *283*, 588–590. [[CrossRef](#)] [[PubMed](#)]
21. Katan, M.B. Apolipoprotein e isoforms, serum cholesterol, and cancer. *Lancet* **1986**, *1*, 507–508. [[CrossRef](#)]
22. Davey Smith, G.; Ebrahim, S. "Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **2003**, *32*, 1–22. [[CrossRef](#)]
23. Zheng, J.; Baird, D.; Borges, M.-C.; Bowden, J.; Hemani, G.; Haycock, P.; Evans, D.M.; Smith, G.D. Recent developments in Mendelian randomization studies. *Curr. Epidemiol. Rep.* **2017**, *4*, 330–345. [[CrossRef](#)] [[PubMed](#)]
24. Swanson, S.A.; Tiemeier, H.; Ikram, M.A.; Hernan, M.A. Nature as a trialist?: Deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology* **2017**, *28*, 653–659. [[CrossRef](#)] [[PubMed](#)]
25. Holmes, M.V.; Ala-Korpela, M.; Smith, G.D. Mendelian randomization in cardiometabolic disease: Challenges in evaluating causality. *Nat. Rev. Cardiol.* **2017**, *14*, 577. [[CrossRef](#)] [[PubMed](#)]
26. Wald, A. The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.* **1940**, *11*, 284–300. [[CrossRef](#)]
27. Lawlor, D.A.; Harbord, R.M.; Sterne, J.A.; Timpson, N.; Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **2008**, *27*, 1133–1163. [[CrossRef](#)] [[PubMed](#)]
28. Burgess, S.; Butterworth, A.; Thompson, S.G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **2013**, *37*, 658–665. [[CrossRef](#)] [[PubMed](#)]
29. Amin, N.; Byrne, E.; Johnson, J.; Chenevix-Trench, G.; Walter, S.; Nolte, I.M.; kConFab, I.; Vink, J.M.; Rawal, R.; Mangino, M. Genome-wide association analysis of coffee drinking suggests association with cyp1a1/cyp1a2 and nrcam. *Mol. Psychiatry* **2012**, *17*, 1116–1129. [[CrossRef](#)] [[PubMed](#)]
30. Cornelis, M.C.; Monda, K.L.; Yu, K.; Paynter, N.; Azzato, E.M.; Bennett, S.N.; Berndt, S.I.; Boerwinkle, E.; Chanock, S.; Chatterjee, N. Genome-wide meta-analysis identifies regions on 7p21 (ahr) and 15q24 (cyp1a2) as determinants of habitual caffeine consumption. *PLoS Genet.* **2011**, *7*, e1002033. [[CrossRef](#)] [[PubMed](#)]
31. Coffee and Caffeine Genetics Consortium; Cornelis, M.C.; Byrne, E.M.; Esko, T.; Nalls, M.A.; Ganna, A.; Paynter, N.; Monda, K.L.; Amin, N.; Fischer, K. Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption. *Mol. Psychiatry* **2015**, *20*, 647–656. [[CrossRef](#)] [[PubMed](#)]
32. Sulem, P.; Gudbjartsson, D.F.; Geller, F.; Prokopenko, I.; Feenstra, B.; Aben, K.K.; Franke, B.; den Heijer, M.; Kovacs, P.; Stumvoll, M. Sequence variants at cyp1a1-cyp1a2 and ahr associate with coffee consumption. *Hum. Mol. Genet.* **2011**, *20*, 2071–2077. [[CrossRef](#)] [[PubMed](#)]
33. Zhong, V.; Kuang, A.; Danning, R.; Kraft, P.; van Dam, R.; Chasman, D.; Cornelis, M.C. A Genome-Wide Association Study of Habitual Bitter and Sweet Beverage Consumption. 2018, submitted for publication.
34. Fredholm, B.B.; Battig, K.; Holmen, J.; Nehlig, A.; Zvartau, E.E. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. *Pharmacol. Rev.* **1999**, *51*, 83–133. [[PubMed](#)]
35. Hindorf, L.; MacArthur, J.; Morales, J.; Junkins, H.; Hall, P.; Klemm, A.; Manolio, T. Catalogue of Published Genome-Wide Association Studies. Available online: <https://www.ebi.ac.uk/gwas/> (accessed on 1 August 2018).
36. Locke, A.E.; Kahali, B.; Berndt, S.I.; Justice, A.E.; Pers, T.H.; Day, F.R.; Powell, C.; Vedantam, S.; Buchkovich, M.L.; Yang, J. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **2015**, *518*, 197–206. [[CrossRef](#)] [[PubMed](#)]

37. Teslovich, T.M.; Musunuru, K.; Smith, A.V.; Edmondson, A.C.; Stylianou, I.M.; Koseki, M.; Pirruccello, J.P.; Ripatti, S.; Chasman, D.I.; Willer, C.J. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **2010**, *466*, 707–713. [[CrossRef](#)] [[PubMed](#)]
38. Manning, A.K.; Hivert, M.F.; Scott, R.A.; Grimsby, J.L.; Bouatia-Naji, N.; Chen, H.; Rybin, D.; Liu, C.T.; Bielak, L.F.; Prokopenko, I. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycaemic traits and insulin resistance. *Nat. Genet.* **2012**, *44*, 659–669. [[CrossRef](#)] [[PubMed](#)]
39. Taylor, A.E.; Davey Smith, G.; Munafò, M.R. Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank. *Addiction* **2018**, *113*, 148–157. [[CrossRef](#)] [[PubMed](#)]
40. McMahon, G.; Taylor, A.E.; Smith, G.D.; Munafò, M.R. Phenotype refinement strengthens the association of *AHR* and *CYP1A1* genotype with caffeine consumption. *PLoS ONE* **2014**, *9*, e103448. [[CrossRef](#)] [[PubMed](#)]
41. Cornelis, M.C.; Kacprowski, T.; Menni, C.; Gustafsson, S.; Pivin, E.; Adamski, J.; Artati, A.; Eap, C.B.; Ehret, G.; Friedrich, N. Genome-wide association study of caffeine metabolites provides new insights to caffeine metabolism and dietary caffeine-consumption behavior. *Hum. Mol. Genet.* **2016**, *25*, 5472–5482. [[PubMed](#)]
42. Davies, N.M.; Holmes, M.V.; Smith, G.D. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ* **2018**, *362*, k601. [[CrossRef](#)] [[PubMed](#)]
43. Nordestgaard, A.T.; Nordestgaard, B.G. Coffee intake, cardiovascular disease and all-cause mortality: Observational and Mendelian randomization analyses in 95,000–223,000 individuals. *Int. J. Epidemiol.* **2016**, *45*, 1938–1952. [[CrossRef](#)] [[PubMed](#)]
44. Nordestgaard, A.T.; Thomsen, M.; Nordestgaard, B.G. Coffee intake and risk of obesity, metabolic syndrome and type 2 diabetes: A Mendelian randomization study. *Int. J. Epidemiol.* **2015**, *44*, 551–565. [[CrossRef](#)] [[PubMed](#)]
45. Kwok, M.K.; Leung, G.M.; Schooling, C.M. Habitual coffee consumption and risk of type 2 diabetes, ischemic heart disease, depression and Alzheimer’s disease: A Mendelian randomization study. *Sci. Rep.* **2016**, *6*, 36500. [[CrossRef](#)] [[PubMed](#)]
46. Treur, J.L.; Taylor, A.E.; Ware, J.J.; Nivard, M.G.; Neale, M.C.; McMahon, G.; Hottenga, J.J.; Baselmans, B.M.; Boomsma, D.I.; Munafò, M.R. Smoking and caffeine consumption: A genetic analysis of their association. *Addict. Biol.* **2017**, *22*, 1090–1102. [[CrossRef](#)] [[PubMed](#)]
47. Taylor, A.E.; Martin, R.M.; Geybels, M.S.; Stanford, J.L.; Shui, I.; Eeles, R.; Easton, D.; Kote-Jarai, Z.; Amin Al Olama, A.; Benlloch, S. Investigating the possible causal role of coffee consumption with prostate cancer risk and progression using Mendelian randomization analysis. *Int. J. Cancer* **2017**, *140*, 322–328. [[CrossRef](#)] [[PubMed](#)]
48. Ware, J.J.; Tanner, J.A.; Taylor, A.E.; Bin, Z.; Haycock, P.; Bowden, J.; Rogers, P.J.; Davey Smith, G.; Tyndale, R.F.; Munafò, M.R. Does coffee consumption impact on heaviness of smoking? *Addiction* **2017**, *112*, 1842–1853. [[CrossRef](#)] [[PubMed](#)]
49. Bjørngaard, J.H.; Nordestgaard, A.T.; Taylor, A.E.; Treur, J.L.; Gabrielsen, M.E.; Munafò, M.R.; Nordestgaard, B.G.; Åsvold, B.O.; Romundstad, P.; Davey Smith, G. Heavier smoking increases coffee consumption: Findings from a Mendelian randomization analysis. *Int. J. Epidemiol.* **2017**, *46*, 1958–1967. [[CrossRef](#)] [[PubMed](#)]
50. Larsson, S.C.; Traylor, M.; Malik, R.; Dichgans, M.; Burgess, S.; Markus, H.S. Modifiable pathways in Alzheimer’s disease: Mendelian randomisation analysis. *BMJ* **2017**, *359*, j5375. [[CrossRef](#)] [[PubMed](#)]
51. Verweij, K.J.; Vinkhuyzen, A.A.; Benyamin, B.; Lynskey, M.T.; Quaye, L.; Agrawal, A.; Gordon, S.D.; Montgomery, G.W.; Madden, P.A.; Heath, A.C. The genetic aetiology of cannabis use initiation: A meta-analysis of genome-wide association studies and a SNP-based heritability estimation. *Addict. Biol.* **2013**, *18*, 846–850. [[CrossRef](#)] [[PubMed](#)]
52. Ong, J.-S.; Hwang, L.-D.; Cuellar-Partida, G.; Martin, N.G.; Chenevix-Trench, G.; Quinn, M.C.; Cornelis, M.C.; Gharahkhani, P.; Webb, P.M.; MacGregor, S. Assessment of moderate coffee consumption and risk of epithelial ovarian cancer: A Mendelian randomization study. *Int. J. Epidemiol.* **2017**, *47*, 450–459. [[CrossRef](#)] [[PubMed](#)]
53. Larsson, S.C.; Carlström, M. Coffee consumption and gout: A Mendelian randomisation study. *Ann. Rheum. Dis.* **2018**, *77*, 1544–1546. [[CrossRef](#)] [[PubMed](#)]
54. Treur, J.L.; Gibson, M.; Taylor, A.E.; Rogers, P.J.; Munafò, M.R. Investigating genetic correlations and causal effects between caffeine consumption and sleep behaviours. *J. Sleep Res.* **2018**, *3*, e12695. [[CrossRef](#)] [[PubMed](#)]

55. Noyce, A.J.; Kia, D.; Heilbron, K.; Jepson, J.; Hemani, G.; Hinds, D.; Lawlor, D.A.; Smith, G.D.; Hardy, J.; Singleton, A. Tendency towards being a “morning person” increases risk of Parkinson’s disease: Evidence from mendelian randomisation. *bioRxiv* **2018**. [[CrossRef](#)]
56. Zhou, A.; Taylor, A.E.; Karhunen, V.; Zhan, Y.; Rovio, S.P.; Lahti, J.; Sjögren, P.; Byberg, L.; Lyall, D.M.; Auvinen, J. Habitual coffee consumption and cognitive function: A Mendelian randomization meta-analysis in up to 415,530 participants. *Sci. Rep.* **2018**, *8*, 7526. [[CrossRef](#)] [[PubMed](#)]
57. Lee, Y.H. Investigating the possible causal association of coffee consumption with osteoarthritis risk using a Mendelian randomization analysis. *Clin. Rheumatol.* **2018**, in press. [[CrossRef](#)] [[PubMed](#)]
58. Burgess, S.; Thompson, S.G. Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **2015**, *181*, 251–260. [[CrossRef](#)] [[PubMed](#)]
59. Hemani, G.; Bowden, J.; Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **2018**, *27*, R195–R208. [[CrossRef](#)] [[PubMed](#)]
60. Burgess, S.; Bowden, J.; Fall, T.; Ingelsson, E.; Thompson, S.G. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* **2017**, *28*, 30. [[CrossRef](#)] [[PubMed](#)]
61. Burgess, S.; Dudbridge, F.; Thompson, S.G. Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Stat. Med.* **2016**, *35*, 1880–1906. [[CrossRef](#)] [[PubMed](#)]
62. Bowden, J.; Davey Smith, G.; Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* **2015**, *44*, 512–525. [[CrossRef](#)] [[PubMed](#)]
63. Bowden, J.; Davey Smith, G.; Haycock, P.C.; Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **2016**, *40*, 304–314. [[CrossRef](#)] [[PubMed](#)]
64. Verbanck, M.; Chen, C.-Y.; Neale, B.; Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **2018**, *50*, 693–698. [[CrossRef](#)] [[PubMed](#)]
65. Zhu, Z.; Zheng, Z.; Zhang, F.; Wu, Y.; Trzaskowski, M.; Maier, R.; Robinson, M.R.; McGrath, J.J.; Visscher, P.M.; Wray, N.R. Causal associations between risk factors and common diseases inferred from gwas summary data. *Nat. Commun.* **2018**, *9*, 224. [[CrossRef](#)] [[PubMed](#)]
66. Cho, Y.; Shin, S.-Y.; Won, S.; Relton, C.L.; Smith, G.D.; Shin, M.-J. Alcohol intake and cardiovascular risk factors: A Mendelian randomisation study. *Sci. Rep.* **2015**, *5*, 18422. [[CrossRef](#)] [[PubMed](#)]
67. Munafò, M.R.; Tilling, K.; Taylor, A.E.; Evans, D.M.; Davey Smith, G. Collider scope: When selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **2017**, *47*, 226–235. [[CrossRef](#)] [[PubMed](#)]
68. Paternoster, L.; Tilling, K.; Smith, G.D. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLoS Genet.* **2017**, *13*, e1006944. [[CrossRef](#)] [[PubMed](#)]
69. Glymour, M.M.; Tchetgen, E.J.; Robins, J.M. Credible Mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions. *Am. J. Epidemiol.* **2012**, *175*, 332–339. [[CrossRef](#)] [[PubMed](#)]
70. Kendler, K.S.; Schmitt, E.; Aggen, S.H.; Prescott, C.A. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. *Arch. Gen. Psychiatry* **2008**, *65*, 674–682. [[CrossRef](#)] [[PubMed](#)]
71. Vanyukov, M.M.; Tarter, R.E.; Kirillova, G.P.; Kirisci, L.; Reynolds, M.D.; Kreek, M.J.; Conway, K.P.; Maher, B.S.; Iacono, W.G.; Bierut, L. Common liability to addiction and “gateway hypothesis”: Theoretical, empirical and evolutionary perspective. *Drug Alcohol Depend.* **2012**, *123* (Suppl. 1), S3–S17. [[CrossRef](#)]
72. Haycock, P.C.; Burgess, S.; Wade, K.H.; Bowden, J.; Relton, C.; Davey Smith, G. Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. *Am. J. Clin. Nutr.* **2016**, *103*, 965–978. [[CrossRef](#)] [[PubMed](#)]
73. Relton, C.L.; Davey Smith, G. Two-step epigenetic Mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* **2012**, *41*, 161–176. [[CrossRef](#)] [[PubMed](#)]

74. Burgess, S.; Daniel, R.M.; Butterworth, A.S.; Thompson, S.G.; Consortium, E.-I. Network Mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int. J. Epidemiol.* **2014**, *44*, 484–495. [[CrossRef](#)] [[PubMed](#)]
75. Sanderson, E.; Smith, G.D.; Windmeijer, F.; Bowden, J. An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings. *bioRxiv* **2018**. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).