

Assessing causal relationships using genetic proxies for exposures: an introduction to Mendelian randomization

Srinivasa Vittal Katikireddi¹ , Michael J. Green¹, Amy E. Taylor^{2,3} , George Davey Smith² & Marcus R. Munafò^{2,3}

MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, UK,¹ MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK² and UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK³

ABSTRACT

Background and aims Studying the consequences of addictive behaviours is challenging, with understanding causal relationships from observational data being particularly difficult. For example, people who smoke or drink excessively are often systematically different from those who do not, are less likely to participate in research and may misreport their behaviours when they do. Furthermore, the direction of causation between an addictive behaviour and outcome may be unclear. Mendelian randomization (MR) offers potential solutions to these problems. **Methods** We describe MR's principles and the criteria under which it is valid. We identify challenges and potential solutions in its application (illustrated using two applied examples) and describe methodological extensions in its application. **Results** MR is subject to certain assumptions, and requires the availability of appropriate genetic data, large sample sizes and careful design and conduct. However, it has already been applied successfully to the addiction literature. The relationship between alcohol consumption (proxied by a variant in the *ADH1B* gene) and cardiovascular risk has been investigated, finding that alcohol consumption increases risk, with no evidence of a cardioprotective effect at moderate consumption levels. In addition, heaviness of smoking (proxied by a variant in the *CHRNA5-A3-B4* gene cluster) and risk of depression and schizophrenia have been investigated, with no evidence of a causal effect of smoking on depression but some evidence of a causal effect on schizophrenia. **Conclusions** Mendelian randomization analyses are already producing robust evidence for addiction-related practice and policy. As genetic variants associated with addictive behaviours are identified, the potential for Mendelian randomization analyses will grow. Methodological developments are also increasing its applicability.

Keywords Addictive behaviour, causality, econometric models, epidemiological methods, genetic epidemiology, Mendelian randomization analysis.

Correspondence to: Srinivasa Vittal Katikireddi, MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Top floor, 200 Renfield Street, Glasgow G2 3 QB, UK. E-mail: vittal.katikireddi@glasgow.ac.uk

Submitted 11 March 2017; initial review completed 12 April 2017; final version accepted 11 September 2017

INTRODUCTION

Determining whether associations are causal is central to much addiction research but is challenging, with many observational associations unlikely to reflect causal relationships [1]. Randomized controlled trials (RCTs), which support stronger causal inference, are not suited to all research questions—particularly as their external validity may be limited [2–4]. Randomizing long-term behaviours or environmental exposures in humans is unethical and impractical. Many causal questions, such as the long-term consequences of consuming potentially harmful, addictive substances, cannot be answered with RCTs.

Mendelian randomization (MR) provides a tool for assessing the causal effects of behaviours on outcomes, although only when genetic variants associated with behaviours are known [5–8]. While previous reviews of MR exist [9], here we provide an up-to-date general introduction targeted specifically at addiction researchers. We note that other approaches to causal inference using observational data exist (including natural experiment approaches and statistical techniques such as propensity score-matching, time-series analysis and structural equation modelling) [10,11]. We start by revisiting challenges to causal inference in traditional observational studies, explain how MR studies potentially overcome them and outline challenges and possible solutions when applying MR. Throughout,

we illustrate MR's principles with two case studies: tobacco smoking as a possible cause of mental health problems (Box 2) and alcohol consumption as a possible cause of cardiovascular disease (Box 3). We conclude with some emerging methodological developments.

CHALLENGES TO CAUSAL INFERENCE

Traditional observational studies face three major threats to establishing whether or not an association is causal [12].

First, characteristics of people with addictive behaviours (e.g. those who smoke or drink excessively) may differ systematically from other people. In naive comparisons of exposed and unexposed groups, these confounding factors are often responsible for observed differences in outcomes. Theoretically, if all confounding factors were measured and accounted for perfectly, an observational study could establish the effect of a behaviour accurately (provided other biases did not exist) [13]. However, in practice it is difficult or impossible to identify all potential confounders. Furthermore, adequate control of confounding during statistical analysis requires accurate measurement of confounders, with even modest measurement error resulting in bias [14].

Secondly, it can be difficult to establish the direction of causation (i.e. whether reverse causation exists). While longitudinal studies may help (and are therefore more useful in causal inference), this is not always the case; the timing of the outcome in relation to behaviour may be uncertain. For example, when

examining alcohol consumption and heart disease (Box 2) it is possible that behaviour change occurred before the early stages of disease were detected and diagnosed.

Thirdly, collider bias may occur, whereby stratification on a common effect can result in a spurious correlation between otherwise independent variables [12]. Figure 1a illustrates this principle (Fig. 1b–1d illustrates collider bias in the context of MR, and is discussed later). This threat to causal inference is perhaps less intuitive than either of the above, but can impact upon the strength and direction of associations observed [12,15,16]. For example, moderate alcohol drinkers are more likely to participate in research [17–19]. If those without cardiovascular disease (CVD) are also more likely to participate, estimates of association between alcohol consumption and CVD in observed populations will be biased. Figure 2 demonstrates this with simple, hypothetical data. Limiting the analysis to those who participate constitutes conditioning on a common effect and may induce bias [20]: if an individual in the study sample is a heavy drinker, then they will be less likely to have CVD. Conditioning on other variables (e.g. as a result of stratification or statistical adjustment) can, similarly, result in bias.

PRINCIPLES OF MENDELIAN RANDOMIZATION

In MR, genetic variants are used as proxies for the exposure of interest, which helps to avoid some of the problems

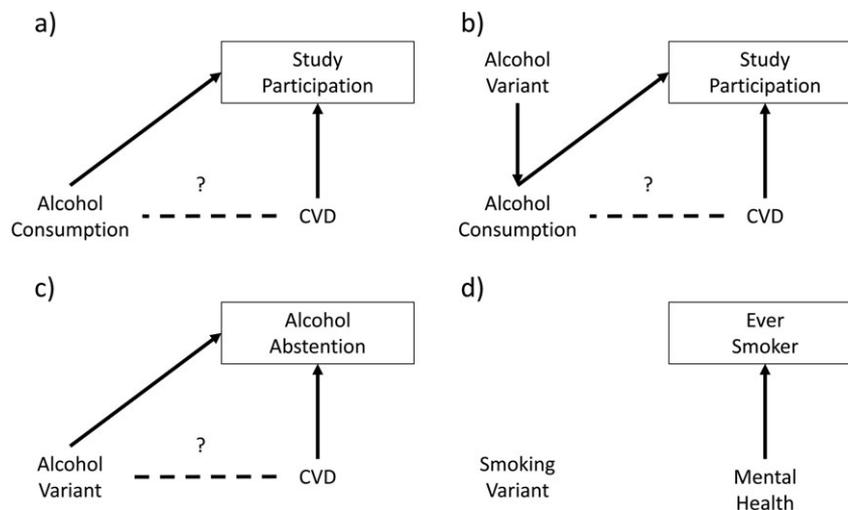


Figure 1 Directed acyclic graphs illustrating collider bias. (a) Collider bias within a traditional observational study arising from sample selection. The box around study participation indicates stratification on this variable. As study participation is influenced by both alcohol consumption and cardiovascular disease (CVD), stratification induces an association between these variables, indicated by the dashed line. (b) Mendelian randomization study which is still subject to collider bias. (c) Abstinence is influenced by the genetic variant and the health outcome, inducing collider bias when stratifying by abstinence. (d) The genetic variant is not associated with whether someone ever becomes a smoker, so there is no collider bias when stratifying on smoking status

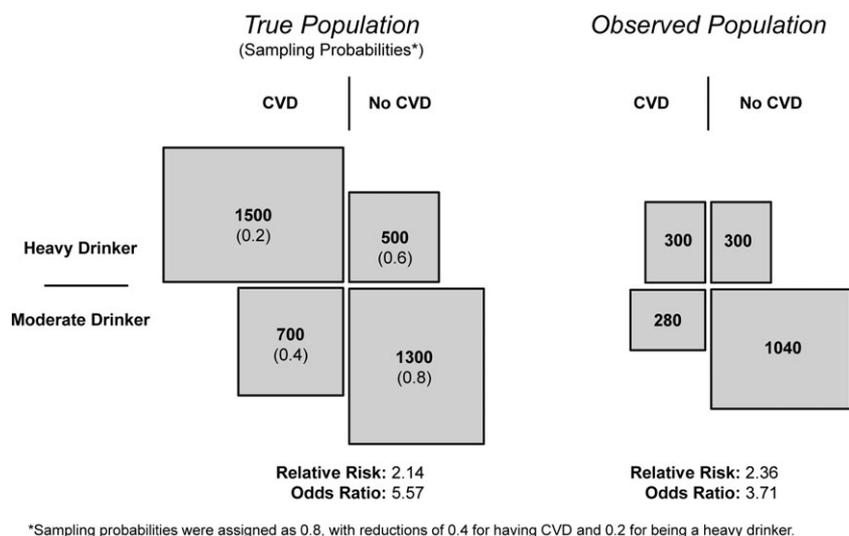


Figure 2 An illustration of collider bias arising from sample selection in a hypothetical study investigating the effect of alcohol consumption on cardiovascular disease (CVD). The left-hand side of the figure demonstrates the relationship between a binary measure of drinking status and cardiovascular disease (based on fictitious data). Numbers in brackets indicate the probability of being recruited into the observed study population on the right-hand side. Differences between the relative risks and odds ratios illustrate the collider bias arising from the selection process

described above. MR is an example of instrumental variable (IV) analysis that has long been used by economists to study causal effects [21]. Randomization in an experimental study might be considered the purest form of instrumental variable [22]. An instrumental variable is a proxy for the exposure of interest. While the correlation between instrument and proxy does not need to be strong, a poor correlation can be problematic and is referred to as a weak instrument (discussed later). The instrument should be unrelated to confounders and should impact the outcome only through its effect on the exposure. In an unbiased RCT, allocation by randomization is associated with the exposure group and is independent of confounders, so that the only pathway between allocation group and the outcome is through the exposure (treatment). An example of instrumental variable analysis using observational data from the economics literature is the use of minimum legal drinking age within US states as an instrument to study the effect of youth drinking on health and social outcomes [23]. Assumptions underpinning this approach include observed and unobserved state characteristics that influence youth drinking being uncorrelated with their minimum legal drinking age policy, and that associations between the policy and outcomes operate only via youth drinking. Both these assumptions can be questioned (e.g. the assumption of the instrument being independent of confounders may be invalid if states that reduced legal drinking ages earlier experienced greater alcohol-related harm previously).

In MR, genetic variants can be used as IVs to either assess whether a causal effect between exposure and outcome exists at all, or to measure the magnitude of the causal effect of the exposure on the outcome. Box 1 provides an overview of the principles for conducting a MR

analysis. For a genetic variant to be a valid IV, it must satisfy three conditions:

- 1 The genetic variant must be associated with the exposure of interest (i.e. the behaviour being studied),
- 2 The genetic variant must be independent of any confounders of the exposure–outcome relationship being studied; and
- 3 The genetic variant should only affect the outcome through the exposure of interest.

Box 1: Conducting a Mendelian randomization analysis

The following is intended as a general guide to the principal stages of an MR analysis, acknowledging that this (and particularly step 7) may evolve as newer techniques emerge. Interested readers may wish to consider three other reviews [24–26]. We present the examples in Boxes 2 and 3 structured around these steps so that the reader can see how the techniques are applied in practice.

- 1 *Define the research question, objectives and protocol:* define exposure(s) and outcome(s), data analysis methods, variables to be used, statistical power calculations [27], etc. This would typically include identification of a genetic variant that is known to be robustly associated with the exposure (or behaviour) of interest.
- 2 *Identify data source(s):* identify potential data sources [e.g. published reports, summary results from genome-wide association study (GWAS) consortia or individual-level data] for the association between

the genetic variant and the exposure, and for the association between the genetic variant and outcome. Two-sample MR (see below) requires data for both associations to come from the different sources (see [68] for more details). In one sample MR, the gene exposure and gene outcome associations are estimated within the same sample.

- 3 *Estimate the gene–exposure association:* (i.e. test condition 1) for example, by regression of the exposure variable on the effect allele (or genetic risk score if multiple genetic variants are in use). If possible, calculate a partial *F*-statistic, which provides an indication of the strength of the genetic instrument [28].
- 4 *Estimate associations between the genetic variant and measured confounders:* (i.e. a falsification test of condition 2) again, this might take the form of regression of measured confounders on the effect allele or genetic risk score. If multiple study populations are in use, do this for all populations where data on measured confounders are available.
- 5 *Estimate the gene–outcome association:* for example, by regression of the outcome on the effect allele or genetic risk score, this is often referred to as the ‘reduced form’ [29,30].
- 6 *Estimate the magnitude of the causal effect:* this step is not necessary if the researcher is interested only in whether a causal effect exists, but if there is interest in the magnitude of effect this can be estimated as the ratio of the gene–outcome (reduced form) to the association between the exposure (or behaviour) and the genetic variant. Alternatively, one can first estimate the relationship between the genetic

variant and the behaviour—for example, using conventional regression methods—and then estimate the relationship between the predicted behaviour from the first regression model and the outcome. In practice, rather than estimating two separate models, a jointly estimated two-stage regression model may be used to take account of uncertainty in the predicted values from the first stage. It is worth noting that the magnitude of the causal effect should not be calculated for certain genetic instruments (e.g. *CHRNA5-A3-B4* which is used in Box 2 cannot be used to assess the magnitude of the causal effect for cigarettes per day on lung cancer) [31].

- 7 *Assess the plausibility of assumptions:* consider whether in the study setting the results could have been affected by pleiotropy, canalization, population stratification or unmeasured confounding (see section on challenges with MR and strategies to overcome them). These issues should at least be discussed in reporting results or where possible, formally tested.

These conditions are illustrated graphically in Fig. 3. Figure 3a illustrates a scenario where all conditions have been met. Figure 3b–d shows violations of the three conditions.

Figure 3b shows violation of the first condition, that the genetic variant is associated with the behaviour of interest. Genetic variants often do not act as direct proxies. For example, the rs1051730 variant in the *CHRNA5-A3-B4* nicotinic receptor subunit gene cluster is associated with the heaviness of smoking in smokers, rather than smoking

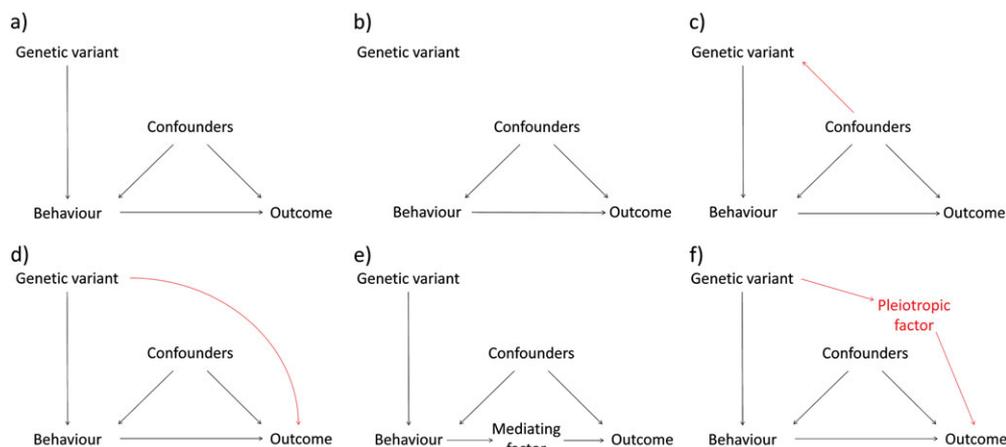


Figure 3 Directed acyclic graphs illustrating the assumptions underpinning valid Mendelian randomization (MR) studies. (a) All three assumptions for valid analysis are met. (b) No relationship between genetic variant and exposure, therefore assumption 1 is not met. (c) The genetic variant is not independent of confounders, therefore assumption 2 is not met. (d) The genetic variant does not exert its effect on the outcome only through the behaviour of interest, therefore assumption 3 (the ‘exclusion restriction’) is not met. (e) Mediated (or vertical) pleiotropy, where the behaviour of interest exerts its impact on the outcome via other intermediate factors. MR remains valid in this situation. (f) Biological (or horizontal) pleiotropy, where the genetic variant exerts effects on the outcome via both the behaviour of interest and via another pleiotropic factor. Note that this is an example of assumption 3 not being met. [Colour figure can be viewed at wileyonlinelibrary.com]

uptake per se (see Box 2 for an example using this variant for MR). It is therefore important to consider what specific aspect of behaviour the genetic variant reflects when interpreting MR results.

Box 2: Cigarette smoking and mental health

Define the research question, objectives and protocol: prevalence of smoking is higher among individuals with psychiatric disease (e.g. depression, schizophrenia) than in the general population, and individuals with these conditions tend to smoke more heavily [32]. However, it is unclear whether the relationship between smoking and mental health is causal and, if so, what the direction of causality is. Individuals may smoke in order to relieve the symptoms of psychiatric disease, but it is also possible that smoking could increase risk of psychiatric disease. Smoking heaviness was the exposure of interest and two mental health outcomes were considered: depression and schizophrenia, measured both as diagnoses and by medication use. The rs1051730 variant in the *CHRNA5-A3-B4* cluster of nicotinic receptor subunit genes is associated with increased smoking heaviness (number of cigarettes smoked per day) among smokers, so was considered a potential instrumental variable. Associations were estimated using traditional regression.

Identify data source: individual-level data from 63 296 participants in the Copenhagen General Population Study (CPGS).

Estimate the gene–exposure association: in the CGPS study, ever smokers with 0.1 and 2 risk (T) alleles smoked an average of 13.6, 14.5 and 15.6 cigarettes per day, respectively (P -value = 1×10^{-47}).

Estimate associations between the genetic variant and measured confounders: there was no clear evidence that the rs1051730 risk variant was associated with age, sex, education, marital status, income, alcohol consumption or physical activity.

Estimate the gene–outcome association: among ever smokers, the smoking heaviness increasing (T) allele of rs1051730 was associated with increased odds of antipsychotic medication use [odds ratio (OR) for TT homozygote compared to CC homozygote: 1.16 (95% confidence interval (CI) = 1.02–1.31]. Due to the nature of the allele used for the MR analysis, it is not possible to estimate a useful causal effect estimate for the number of cigarettes per day on the outcome. A similar trend was observed for schizophrenia diagnoses,

although statistical evidence for this association was weak. There was little evidence that this variant was associated with depression or antidepressant medication use among ever smokers.

Assess the plausibility of assumptions: the rs1051730 variant is associated closely with the *CHRNA5* nicotinic receptor subunit gene, which has been shown to alter response to nicotine and subsequently affects how much tobacco is consumed among smokers. Therefore, there is a plausible biological mechanism linking this variant with smoking behaviour. It is believed not to influence the likelihood of someone becoming a smoker in the first place (i.e. Fig. 1d), so there is little risk of collider bias when stratifying by smoking status. Analyses stratified by smoking status showed that among never smokers the rs1051730 T allele was not associated with antipsychotic medication use (OR for TT homozygote compared to CC homozygote: 1.07 (95% CI = 0.87–1.31), which provides some evidence against pleiotropy because rs1051730 cannot be associated with heaviness of smoking in never smoking individuals. However, this effect estimate was also not clearly different from the effect estimate among smokers, so the results should be treated with some caution.

The second condition ensures that the genetic variant is not related to confounders (violated in Fig. 3c). While condition 1 can be tested empirically, condition 2 can, at best, only be tested partially. Researchers may demonstrate statistical independence of the genetic variant from measured confounders (for example in Box 2 the rs1051730 variant was not associated with several measured confounders). However, a theoretically informed argument for why an instrument should be independent of other unmeasured confounders is also needed [33]. There are good reasons to expect this with genetic variants. According to Mendel's first law, each of the parent's two copies of a given section of DNA has an equal chance of being inherited, with the environment not influencing which copy produces a viable fertilized egg [34]. The second law states that the two copies of a gene are inherited approximately independently from each other and from other genetic variants. The most common version of a specific genetic variant is referred to as the major allele and the least common is called the minor allele. These two laws in combination imply that any specific allele should be distributed randomly across the population, provided it has been transmitted stably across several generations and people's choice of partner is not influenced by the allele (i.e. there is no assortative mating).

Within econometrics, condition 3 is referred to as the 'exclusion restriction'. This means that the causal pathway

between the genetic instrument and the outcome occurs only through the behaviour of interest. It would be violated if the genetic variant had an impact upon the outcome through factors other than the exposure of interest—i.e. the genetic variant was acting as a proxy for other factors as well as the behaviour being studied (see Fig. 3d). Condition 3 also cannot be tested directly, but indirect supportive evidence can help in its evaluation (discussed later).

When a genetic variant is a valid IV [35], MR can overcome the first two of the three threats to causal inference outlined above. First, with regard to confounding, genetic variants are expected to be distributed approximately randomly across the population, and should therefore be independent of confounders of exposure–outcome relationships [1]. Secondly, genotype is determined at conception and germline DNA is not modified thereafter, precluding reverse causation. With respect to the third threat, it has been suggested that selection bias may be less problematic in MR studies than in observational studies; for example, the distribution of genetic variants was found to be similar in blood donors (a highly selected group) compared with the UK general population [36,37]. However, collider bias may still be induced if both the behaviour and outcome of interest are related to study participation (see Fig. 1b). Even weak selection biases may influence outcomes of MR studies [20], as small biases in estimated gene–outcome associations can result in large changes to the causal behavioural–outcome estimates [38,39]. Thus, compared to conventional observational research, MR overcomes challenges of confounding and reverse causation, but selection bias may remain problematic. MR approaches can also be used to investigate whether behaviours impact upon prognosis after disease diagnosis—for example, while smoking causes lung cancer it is not clear whether it also influences prognosis [40]. MR can therefore be used to understand the effects of addictive behaviours on disease prognosis, which may differ from their effects on aetiology, in turn informing clinical advice given to patients who have received a diagnosis. However, data sources for MR of progression are currently more limited than for studying aetiology.

CHALLENGES WITH MENDELIAN RANDOMIZATION AND STRATEGIES TO OVERCOME THEM

We now describe some of the most important challenges with MR and commonly used strategies for addressing them. At the outset, we note that the need to have genetic data available is an important potential limitation—if DNA has not already been collected it may not be feasible to do so.

First, the genetic variant should be distributed randomly across the sample being analysed. However, the variant's distribution may differ between historically separate human populations, even though it is distributed randomly within each population [41]. If outcome risks also differ between such populations, then the effect estimate may be confounded—a phenomenon referred to as 'population stratification'. Restricting analysis to a single ethnic group (e.g. analysis of European ancestry individuals only) or statistical adjustment for ancestral information can reduce this risk.

Secondly, a crucial challenge is finding an appropriate genetic instrument, typically from a GWAS of the exposure of interest [42]. These studies compare the frequency of genetic variants throughout the human genome in people who exhibit a behaviour with those who do not [43,44]. GWAS, which typically combine large sample size, statistical stringency (to account for the large number of statistical tests) and replication in an independent sample, have a good track record in identifying genetic variants reliably (i.e. alleles) associated with behaviours. The genetic variant need not be related causally to the behaviour, as long as they are associated reliably, so MR does not require knowledge about the function of the genetic variant. Nevertheless, causal inference is strengthened with understanding of the biological process. Without understanding how a gene exerts its function, we can be less certain that any effects are caused genuinely by the behaviour (i.e. assumption 3; see discussion of pleiotropy below).

A third challenge is that very large sample sizes are needed in MR because individual genetic variants typically exert small influences on behaviour and exposure measurement is often poor [45]. Thus, studies may have low statistical power, especially for estimating the magnitude of causal effects, which requires estimation of two associations. Without large sample sizes, weakly associated genetic variants may yield estimates biased towards the naive observational association [46–48]. To counter this, several genetic variants that are all associated with a behaviour may be combined to create a polygenic risk score [49]. This could simply be a count of genetic variants that increase the behaviour, but more refined approaches weight the score so that more predictive genetic variants are weighted more strongly [50]. While polygenic risk scores help address the weak instrument problem, the score must satisfy the three conditions for a valid instrument and may not do so if individual components do not themselves satisfy the three conditions [32].

An additional approach that makes achieving large sample sizes easier is two-sample MR [51]. As noted earlier, two associations are estimated typically when assessing the magnitude of a causal effect in MR: the association between genetic variant and behaviour and the association between

genetic variant and outcome. In two-sample MR these two associations are estimated from different samples, which is unproblematic provided that the underlying population from which the samples are drawn are the same. This reduces the need to access individual participant data.

A fourth challenge is especially problematic when assessing the magnitude (rather than presence) of a causal effect. As noted earlier, condition 3 (the exclusion restriction) requires no causal pathway between the genetic variant and the outcome, except through the exposure of interest. In MR studies, pleiotropy can violate this assumption [52]. Two forms of pleiotropy can be distinguished [53,54]. Mediated (or vertical) pleiotropy occurs when the genetic variant is associated with a factor on the pathway between the behaviour and outcome, but only because of its effect on the behaviour (see Fig. 3e). This does not violate the exclusion restriction, as it is part of the pathway through which the behaviour exerts an effect. In contrast, biological (or horizontal) pleiotropy occurs when the genetic variant impacts upon a different biological pathway unrelated to the behaviour of interest, and therefore violates condition 3 (Fig. 3f). Methods to investigate, and to some extent relax, this assumption are described later.

Finally, cautious interpretation is necessary, bearing in mind the nature of the causal effect being estimated. For example, the study in Box 3 provides an estimate of the effect of the rs1229984 genetic variant on heart disease, indicating a causal effect of alcohol consumption. However, this genetic variant was associated both with units of alcohol consumed per week and with binge drinking. Attempts to estimate the magnitude of causal effects for specific drinking behaviour patterns, such as units per week, or binge drinking would be biased due to violation of condition 3 (as the genetic variant affects both behaviour patterns). Even when a genetic variant meets the three conditions to be a valid instrument, it may not provide an unbiased estimate of the effect of the specific aspect of behaviour that is of interest or has been measured [55]. Also, as genetic variants are established at conception, the estimated causal effect is of a life-long tendency towards a certain behaviour (or susceptibility to its impact) [56,57]. Furthermore, physiological adaptations may reduce a gene's effects (referred to as canalization) [58], while a gene's effects may be observed only under specific environments or exert impacts only at specific points during the life course (i.e. critical periods).

Box 3: Alcohol consumption and coronary heart disease

Define the research question, objectives and protocol: observational studies often report a J-shaped association between alcohol and cardiovascular

outcomes such as coronary heart disease [59]. Risk is lowest for light to moderate drinkers, but increases for non-drinkers and for heavier or more hazardous drinkers. Increased risk among heavy drinkers and abstainers may be due to reverse causation (e.g. abstainers may not drink because they have poorer health), selection bias or confounding from other social, life-style or health factors. Alcohol consumption, measured in weekly units (where 1 unit = 4 g ethanol), was the main exposure of interest. The primary outcome considered was coronary heart disease. The minor allele at rs1229984 of the alcohol dehydrogenase 1b gene (*ADH1B*) has been identified in previous research as being associated with lower alcohol consumption and so was identified as a potential genetic instrument. Associations were estimated with standard regression techniques in multiple studies and then pooled using meta-analysis.

Identify data sources: individual participant data on 261 991 participants of European ancestry were gathered from across 56 genetic studies.

Estimate the gene–exposure association: carriers of the risk allele at rs1229984 consumed fewer units of alcohol per week (17.2% fewer on average; 95% CI = 18.9–15.6%), had lower odds of heavy drinking (OR = 0.70; 95% CI = 0.68–0.73) or binge drinking (OR = 0.78; 95% CI = 0.73–0.84) and higher odds of abstinence (OR = 1.27; 95% CI = 1.21–1.34) compared to non-carriers.

Estimate associations between the genetic variant and measured confounders: the rs1229984 risk variant was not associated with physical activity or with most measures of smoking, although it was associated with slightly higher odds of ever smoking (OR = 1.06; 95% CI = 1.02–1.09), which is in the opposite direction to observational studies, and with slightly more years of education (0.04 standard deviations; 95% CI = 0.01–0.08).

Estimate the gene–outcome association: the rs1229984 risk variant was associated with reduced odds of heart disease among the whole sample (OR = 0.90; 95% CI = 0.84–0.96) and among drinkers (OR = 0.86; 95% CI = 0.78–0.94), suggesting that lower consumption is protective. Associations of the risk allele with heart disease did not differ further across light, moderate and heavy drinkers. This is contrary to the increased risk that would have been expected among light/moderate drinkers if moderate alcohol consumption had a true, causal protective effect relative to very low consumption. This suggests a linear, rather than a

J-shaped relationship between alcohol consumption and heart disease.

Assess the plausibility of assumptions: the *ADH1B* gene is involved in metabolizing alcohol, with the rs1229984 risk variant increasing the likelihood of unpleasant symptoms after consumption. Thus, there is a plausible biological mechanism connecting the genetic variant to alcohol consumption. The weak associations between the genetic variant and measured confounders are unlikely to explain any association between the genetic variant and heart disease, but it is possible that unmeasured confounders could be biasing results if transmission of the genetic variant across generations is related to alcohol consumption (i.e. assortative mating is occurring). Among non-drinkers the rs1229984 risk variant was not associated with heart disease (OR = 0.98; 95% CI = 0.88–1.10), suggesting that it has no effect other than via reductions in alcohol consumption. This estimate may have been affected by collider bias because the genetic variant is associated with both abstinence and heaviness of drinking among drinkers (see Fig. 1c), but there were no associations with measured confounders in stratified analyses, suggesting that these factors were unlikely to be biasing results.

Implications for policy and practice: light alcohol consumption does not protect from ischaemic heart disease, therefore efforts to reduce alcohol consumption are not likely to have any adverse impact on cardiovascular risk prevention.

ASSESSING THE ROBUSTNESS OF MENDELIAN RANDOMIZATION STUDIES

As noted above, biological pleiotrophic effects (violating assumption 3) threaten the validity of MR. While there remain no definitive ways of addressing this, several approaches now exist to explore potential bias. A theory-informed approach is to make potentially informative comparisons to check the plausibility of assumption 3. For example, looking throughout countries with differing cultural norms for alcohol consumption can help to establish whether biological pleiotropy exists. In East Asian countries women tend not to consume alcohol, so genetic variants related to alcohol consumption would be expected to be associated with alcohol-related disease outcomes in men but not in women [60]. Cho and colleagues confirmed this by fitting a statistical interaction between the genetic variant and sex when carrying out a MR analysis using a South Korean sample, thereby providing further evidence that alcohol is related causally

to an adverse cardiovascular risk profile [61]. Similarly, in Box 3 a lack of an association between the genetic variant and heart disease among non-drinkers increases confidence that the association among drinkers is due to drinking. Knowledge about the biological function by which the genetic variant exerts an effect is very helpful, as it provides greater confidence that the effect is via the behaviour rather than another mechanism [62]. Box 2 illustrates how biological understanding can inform assessments of whether condition 3 is met in the case of a genetic variant associated with smoking.

Newly developed methods allow empirical investigation of MR assumptions. The funnel plot is used to identify small study bias in systematic reviews by looking for an association between study precision and effect size [63]. In RCTs, increasing sample size should result in less variation in observed effect sizes (producing a symmetrical funnel-shaped plot). The same principle has been applied to MR studies that make use of multiple genetic variants with differing strengths of association with the behaviour of interest [64]. Stronger associations between genetic variants and the exposure should result in less variation and a symmetrical funnel plot. Asymmetry in the funnel plot indicates that the MR assumptions are not met (see Fig. 3). MR Egger regression builds on this to allow for overall biological (horizontal) pleiotropy across multiple genetic instruments to be estimated and the causal effect to be appropriately adjusted for. It can be applied when conducting two-sample MR [65]. An additional sensitivity analysis is to use a weighted median estimator, as the median estimate across several genetic instruments should be less prone to bias from confounders (i.e. violations of assumption 2), provided that a majority of the weighted analysis is based on valid instruments [66]. The assumptions underpinning each of these techniques differ—therefore a consistent pattern of findings strengthens causal inference. However, these techniques require multiple genetic instruments acting as proxies for the same behaviour.

EXTENSIONS TO MENDELIAN RANDOMIZATION

Analytical tools for MR research are developing and being refined rapidly. Bidirectional MR is an extension of the traditional design utilizing genetic markers for different but inter-related outcomes to investigate the direction and magnitude of the causal effects. For example, the causal relationship between cannabis use and schizophrenia remains controversial. Gage and colleagues studied two sets of genetic variants, one related to cannabis initiation and one related to schizophrenia risk, to understand more clearly the direction of causation [67]. They found that schizophrenia-related genetic

variants were related strongly to cannabis initiation while genetic variants linked to cannabis initiation were associated weakly with schizophrenia. The authors therefore concluded that 'cannabis initiation increases the risk of schizophrenia, but the size of the causal effect is small' and there is 'stronger evidence that schizophrenia risk predicts cannabis initiation'. The use of multiple genetic instruments to investigate the direction of causality can be extended to investigate multiple mediating factors [68]. In such 'network' MR, genetic instruments for each mediating factor to be investigated are required, and the genetic instruments must be independent of each other [69].

Finally, factorial MR allows combinations of multiple behaviours to be investigated. Factorial MR is akin to a factorial RCT where the population is in effect allocated randomly to receive any combination of the behaviours under consideration [70]. For example, the combination of excess alcohol consumption and obesity are known to result in far greater risk of liver cirrhosis than would be expected based on their additive effects (i.e. they show evidence of effect modification) [71,72]. However, observational studies of effect modification face threats to causal inference. Factorial MR therefore has similar advantages to a factorial RCT—allowing multiple behaviours to be investigated and compared against each other, singly or in combination. Again, availability of genetic instruments for the different behaviours and their biological independence are important considerations.

CONCLUSIONS

MR adds to the range of study designs available to understand the causal effects of behaviours on outcomes of interest. It helps address key limitations of traditional observational studies, including confounding and reverse causation, but selection bias could remain problematic. MR studies potentially allow researchers to produce more robust evidence on questions of immense relevance to policy and practice. They can provide strong evidence of causation, subject to necessary assumptions which benefit from an understanding of the underpinning biological processes. However, two of the three assumptions underpinning MR cannot be tested definitively. Furthermore, genetic variants known to be associated with behaviours of interest are required and genetic data from a large number of people, including those exhibiting the behaviour of interest, are needed. A range of other causal approaches to observational research are available, with differing underpinning assumptions; their use in combination can be particularly powerful [10,73]. We have provided a broad overview of the topic so that interested readers are able to read critically and interpret findings from MR studies. The use of genetic instruments for gaining causal understanding is already

yielding important insights into addiction research and will probably advance the field substantially in the future.

Declaration of interests

None.

Acknowledgements

S.V.K. is funded by a NRS Scottish Senior Clinical Fellowship (SCAF/15/02). M.G. and S.V.K. receive funding from the Medical Research Council (MC_UU_12017/13 and MC_UU_12017/15) and the Scottish Government Chief Scientist Office (SPHSU13 and SPHSU15). A.E.T., G.D.S. and M.R.M. receive funding from the Medical Research Council (MC_UU_12013/1, MC_UU_12013/6). A.E.T. and M.R.M. are members of the UK Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

References

1. Davey Smith G., Lawlor D. A., Harbord R., Timpson N., Day I., Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007; **4**: e352.
2. Torgerson D. J., Torgerson C. J. *Designing Randomised Trials in Health, Education and the Social Sciences: an Introduction*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan; 2008.
3. Petticrew M., McKee M., Lock K., Green J., Phillips G. In search of social equipoise. *BMJ* 2013; **347**: f4016. <https://doi.org/10.1136/bmj.f4016>.
4. Victora C. G., Habicht J.-P., Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004; **94**: 400–5.
5. Sheehan N. A., Didelez V., Burton P. R., Tobin M. D. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008; **5**: e177.
6. Davey Smith G., Ebrahim S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ* 2005; **330**: 1076–9.
7. Taylor A. E., Ware J. J., Gage S. H., Smith G. D., Munafò M. R. Using molecular genetic information to infer causality in observational data: Mendelian randomisation. *Curr Opin Behav Sci* 2015; **2**: 39–45.
8. Davey Smith G., Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**: 1–22.
9. Didelez V., Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007; **16**: 309–30.
10. Craig P., Katikireddi S. V., Leyland A. H., Popham F. Natural experiments: an overview of methods, approaches and contribution to public health intervention research. *Annu Rev Public Health* 2017; **38**: 39–56.

11. Basu S., Meghani A., Siddiqi A. Evaluating the health impact of large-scale public policy changes: classical and novel approaches. *Amu Rev Public Health* 2017; **38**: 351–70.
12. Hernán M. A., Hernández-Díaz S., Robins J. M. A structural approach to selection bias. *Epidemiology* 2004; **15**: 615–25.
13. Greenland S., Morgenstern H. Confounding in health research. *Amu Rev Public Health* 2001; **22**: 189–212.
14. Fewell Z., Davey Smith G., Sterne J. A. C. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol* 2007; **166**: 646–55.
15. Elwert F., Winship C. Endogenous selection bias: the problem of conditioning on a collider variable. *Amu Rev Sociol* 2014; **40**: 31–53.
16. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003; **14**: 300–6.
17. Gorman E., Leyland A. H., McCartney G., White I. R., Katikireddi S. V., Rutherford L. *et al.* Assessing the representativeness of population-sampled health surveys through linkage to administrative data on alcohol-related outcomes. *Am J Epidemiol* 2014; **180**: 941–8.
18. Gray L., McCartney G., White I. R., Katikireddi S. V., Rutherford L., Gorman E. *et al.* Use of record-linkage to handle non-response and improve alcohol consumption estimates in health survey data: a study protocol. *BMJ Open* 2013; **3**: e002647.
19. Gorman E., Leyland A. H., McCartney G., Katikireddi S. V., Rutherford L., Graham L. *et al.* Adjustment for survey non-representativeness using record-linkage: refined estimates of alcohol consumption by deprivation in Scotland. *Addiction* 2017; **112**: 1270–80.
20. Munafò M. R., Tilling K., Taylor A. E., Evans D. M., Davey Smith G. Collider scope: how selection bias can induce spurious associations. *Int J Epidemiol* 2017; <https://doi.org/10.1093/ije/dyx206>.
21. Wooldridge J. *Introductory Econometrics: A Modern Approach*. New Delhi: Nelson Education; 2015.
22. Heckman J. J. Randomization as an instrumental variable. *Rev Econ Stat* 1995; **78**: 336–41.
23. French M. T., Popovici I. That instrument is lousy! In search of agreement when using instrumental variables estimation in substance use research. *Health Econ* 2011; **20**: 127–46.
24. Hartwig F. P., Davies N. M., Hemani G., Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol* 2016; **45**: 1717–26.
25. Boef A. G. C., Dekkers O. M., le Cessie S. Mendelian randomization studies: A review of the approaches used and the quality of reporting. *Int J Epidemiol* 2015; **44**: 496–511.
26. 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**: 965–78.
27. Brion M.-J. A., Shakhbazov K., Visscher P. M. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013; **42**: 1497–501.
28. Burgess S., Small D. S., Thompson S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* 2015; pii: 0962280215597579.
29. Katan M. B. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986; **327**: 507–8.
30. Angrist J., Krueger A. B. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Perspect* 2001; **15**: 69–85.
31. Holmes M. V., Ala-Korpela M., Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017; **14**: 577–90.
32. Wium-Andersen M. K., Ørsted D. D., Nordestgaard B. G. Tobacco smoking is causally associated with antipsychotic medication use and schizophrenia, but not with antidepressant medication use or depression. *Int J Epidemiol* 2015; **44**: 566–77.
33. Swanson S. A., Hernan M. A. Commentary: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013; **24**: 370–4.
34. Laird N. M., Lange C. Principles of Inheritance: Mendel's Laws and Genetic Models. In: *The Fundamentals of Modern Statistical Genetics*. New York: Springer; 2011, pp. 15–30.
35. Davey Smith G. Mendelian randomization for strengthening causal inference in observational studies: application to gene × environment interactions. *Perspect Psychol Sci* 2010; **5**: 527–45.
36. The Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–78.
37. Ebrahim S., Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 2008; **123**: 15–33.
38. Jackson J. W., Swanson S. A. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology* 2015; **26**: 498–504.
39. Davies N. M. Commentary: An even clearer portrait of bias in observational studies? *Epidemiology* 2015; **26**: 505–8.
40. Paternoster L., Tilling K., Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: conceptual and methodological challenges. *PLoS Genet* 2017; **13**: e1006944.
41. Cardon L. R., Palmer L. J. Population stratification and spurious allelic association. *Lancet* 2003; **361**: 598–604.
42. McCarthy M. I., Abecasis G. R., Cardon L. R., Goldstein D. B., Little J., Ioannidis J. P. A. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–69.
43. Hirschhorn J. N., Daly M. J. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6**: 95–108.
44. Bush W. S., Moore J. H. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 2012; **8**: e1002822.
45. Freeman G., Cowling B. J., Schooling C. M. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int J Epidemiol* 2013; **42**: 1157–63.
46. Bound J., Jaeger D. A., Baker R. M. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc* 1995; **90**: 443–50.
47. Burgess S., Thompson S. G. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* 2011; **30**: 1312–23.
48. Davies N. M., von Hinke Kessler Scholder S., Farbacher H., Burgess S., Windmeijer F., Davey Smith G. The many weak instruments problem and Mendelian randomization. *Stat Med* 2015; **34**: 454–68.
49. Burgess S., Thompson S. G. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013; **42**: 1134–44.
50. Pierce B. L., Ahsan H., VanderWeele T. J. Power and instrument strength requirements for Mendelian randomization

- studies using multiple genetic variants. *Int J Epidemiol* 2011; **40**: 740–52.
51. Burgess S., Timpson N. J., Ebrahim S., Davey Smith G. Mendelian randomization: where are we now and where are we going? *Int J Epidemiol* 2015; **44**: 379–88.
 52. Krebs J. E., Lewin B., Goldstein E. S., Kilpatrick S. T. *Lewin's genes XI*. Burlington, MA: Jones & Bartlett Publishers; 2014.
 53. Burgess S., Thompson S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015; **181**: 251–60.
 54. Tyler A. L., Asselbergs F. W., Williams S. M., Moore J. H. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *Bioessays* 2009; **31**: 220–7.
 55. Taylor A. E., Davies N. M., Ware J. J., VanderWeele T., Davey Smith G., Munafò M. R. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Econ Hum Biol* 2014; **13**: 99–106.
 56. Burgess S., Thompson S. G. *Mendelian randomization: methods for using genetic variants in causal estimation*. Boca Raton, Florida: CRC Press; 2015.
 57. Davey Smith G., Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; **33**: 30–42.
 58. Debat V., David P. Mapping phenotypes: Canalization, plasticity and developmental stability. *Trends Ecol Evol* 2001; **16**: 555–61.
 59. Holmes M. V., Dale C. E., Zuccolo L., Silverwood R. J., Guo Y., Ye Z. *et al.* Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ* 2014; **349**: g4164.
 60. Xu L., Jiang C. Q., Cheng K. K., Au Yeung S. L. R., Zhang W. S., Lam T. H. *et al.* Alcohol use and gamma-glutamyltransferase using a Mendelian randomization design in the Guangzhou Biobank Cohort Study. *PLOS ONE* 2015; **10**: e0137790.
 61. Cho Y., Shin S.-Y., Won S., Relton C. L., Davey Smith G., Shin M.-J. Alcohol intake and cardiovascular risk factors: a Mendelian randomisation study. *Sci Rep* 2015; **5**: 18422.
 62. Evans D. M., Davey Smith G. Mendelian randomization: new applications in the coming age of hypothesis-free causality. *Annu Rev Genomics Hum Genet* 2015; **16**: 327–50.
 63. Egger M., Davey Smith G., Schneider M., Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–34.
 64. 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**: 512–25.
 65. Bowden J., Del Greco M. E., Minelli C., Davey Smith G., Sheehan N. A., Thompson J. R. 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**: 1961–74.
 66. 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**: 304–14.
 67. Gage S. H., Jones H. J., Burgess S., Bowden J., Davey Smith G., Zammit S., *et al.* Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. *Psychol Med* 2016; **47**: 971–80.
 68. Burgess S., Daniel R. M., Butterworth a. S., Thompson S. G. and the EPIC-InterAct Consortium Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol* 2015; **44**: 484–95.
 69. Richmond R. C., Hemani G., Tilling K., Davey Smith G., Relton C. L. Challenges and novel approaches for investigating molecular mediation. *Hum Mol Genet* 2016; **25**: R149–56.
 70. Davey Smith G., Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; **23**: R89–R98.
 71. Liu B., Balkwill A., Reeves G., Beral V. Body mass index and risk of liver cirrhosis in middle aged UK women: prospective study. *BMJ* 2010; **340**: c912.
 72. Hart C. L., Morrison D. S., Batty G. D., Mitchell R. J., Davey Smith G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. *BMJ* 2010; **340**: c1240.
 73. Gage S. H., Munafò M. R., Davey Smith G. Causal inference in developmental origins of health and disease (DOHaD) research. *Annu Rev Psychol* 2016; **67**: 567–85.