Chapter 12

Mendelian Randomisation: A Tool for Assessing Causality in Observational Epidemiology

Nuala A. Sheehan, Sha Meng, and Vanessa Didelez

Abstract

Detection and assessment of the effect of a modifiable risk factor on a disease with view to informing public health intervention policies are of fundamental concern in aetiological epidemiology. In order to have solid evidence that such a public health intervention has the desired effect, it is necessary to ascertain that an observed association or correlation between a risk factor and a disease means that the risk factor is *causal* for the disease. Inferring causality from observational data is difficult, typically due to confounding by social, behavioural, or physiological factors which are difficult to control for and particularly difficult to measure accurately. A possible approach to inferring causality when confounding is believed to be present but unobservable, as it may not even be fully understood, is based on the method of instrumental variables and is known under the name of *Mendelian randomisation* if the instrument is a genetic variant. While testing for the presence of a causal effect using this method is generally straightforward, point estimates of such an effect are only obtainable under additional parametric assumptions. This chapter introduces the concept and illustrates the method and its assumptions with simple real-life examples. It concludes with a brief discussion on pitfalls and limitations.

Key words: Causal inference, Instrumental variable, Confounding

1. Introduction

The study of risk factors for disease is central to epidemiological research. Here, we distinguish between prognostic and aetiological research by considering the notion of risk in its original context of studying conditions thought to be *caused* by a particular factor and not in the broader sense of *predicting* the probability of a condition for prognostic purposes. For the latter, all factors associated with the outcome are of interest, regardless of whether they are *causal* or not. For aetiological research, the focus is on assessing the effects of modifiable exposures on disease with view

M. Dawn Teare (ed.), *Genetic Epidemiology*, Methods in Molecular Biology, vol. 713,
DOI 10.1007/978-1-60327-416-6_12, © Springer Science+Business Media, LLC 2011

to informing health intervention policies. It is hence important to verify that an observed association between the exposure and disease of interest indicates a *causal* relationship between the two. Inferring causality from observed associations is problematic as it is not clear which of two correlated variables is the cause, which the effect, or whether the association is due to another unmeasured factor, or *confounder*. Randomised controlled trials (RCTs) provide the accepted solution since they render reverse causation and confounding implausible. However, RCTs are neither ethical nor practical for many exposures of epidemiological interest, such as exercise, alcohol consumption, and diet regimes. From the practical viewpoint, many exposures develop over years so people cannot be randomised easily to "lifetime" exposure and trials attempting to do so are very costly. Moreover, the population of volunteers in a trial is likely to differ considerably from the general population. Thus, we often have to make causal inferences from observational epidemiological studies and, arguably, we actually need to do so when public health interventions are of interest as we require a representative study population (1).

There have been many success stories where evidence from epidemiological studies has informed public health policy and led to health improvements in the general population. These include the links between smoking and increased risk of lung cancer (2), and between maternal folate supplementation and reduced risk of neural tube defects (3, 4) leading to widespread banning of smoking in public places and the mandatory fortification of cereal flour with folic acid in the USA, Canada, and Chile, for example. There have also been many high-profile failures, where reported associations failed to be replicated in follow-up RCTs. For example, the observation that increased beta-carotene intake reduces the risk of smoking-related cancers was not replicated in the subsequent large-scale RCTs (5–7). More recent failures to replicate observational findings in RCTs include the associations between hormone replacement therapy and cardiovascular disease and between oestrogen levels and Alzheimer's disease or dementia.

There are many reasons why an observational study and an RCT could provide contradictory results. Different dose levels, different durations of follow-up and interactions with other risk factors are usually proposed, but they do not fully explain such discrepancies. The most likely reasons are *confounding* by unobserved lifestyle, socioeconomic factors or baseline health status, *reverse causation*, where the presence of disease influences what is thought to be exposure rather than vice versa, and the usual problems of selection or reporting bias. Since only those associations with high observational support are ever likely to be verified in an RCT, we can only presume that many other reported associations are likely to be non-causal (8). Given the tendency of high-profile findings to persist in the medical literature and thus influence

public health and clinical policy long after they have been refuted by RCT evidence (9), it is important to have alternative methods for assessing causality from observational data. Here, we particularly address the case where we have unobserved confounding factors and so cannot adjust our analyses in the usual way.

Mendelian randomisation is an *instrumental variable* (IV) approach to the problem of inferring causality when unobserved confounding is believed to be likely, possibly because the underlying biological processes are not fully understood (1, 10–16). It uses a well-understood genetic variant, known to be associated with the exposure but without direct effect on the disease, as an instrument. The exposure may itself be a phenotype or a genetically influenced behaviour. The fact that genes are assigned randomly at meiosis (given the parental genes) implies that the instrument should be independent of any unobserved confounding between the exposure and the disease, and so we can think of Mendelian randomisation as a natural imitation of a randomised trial although the randomisation is not, of course, perfect. Reverse causation is not an issue here since genes are determined before birth. The basic idea is that there should be no association between the genetic variant and the disease *unless* the considered exposure or phenotype is actually causal for the disease.

2. Mendelian Randomisation

In this section, we introduce a formal framework for causal inference and the core conditions for IV methods with a brief discussion of some of the implications for testing and estimating causal effects in epidemiological applications. We then illustrate the method with some examples.

2.1. Causal Concepts

We first need to formalise how we distinguish between *association* and *causation*. If we say that a variable X is associated with another variable Y, we mean that observing X is informative, or predictive, for Y. The usual conditional probability notation $P(Y = y|X = x)$ describes the distribution of Y, given that we happen to know that $X = x$ has occurred.

We regard causal inference to be about studying the effect of *intervening* in a particular system (17–20). Other causal frameworks are based on counterfactual or potential outcome variables (21) or structural equation models (15, 22). It can be argued that the notion of intervention is implicit to all these formal approaches to causality (23, 24). Specifically, when we say that X *causes* Y, we mean that *manipulating* or *intervening on* X is informative for Y. Ordinary conditional probability notation does not reflect the changes in the distribution of Y when X is *set* to a

particular value. We need some extra notation in order to formally distinguish between association and causation. We use the notation do($X=x$) to represent the intervention of setting X to a value x, as suggested by Pearl (22). The two conditional distributions $P(Y=y|\text{do}(X=x))$ and $P(Y=y|X=x)$ are not necessarily the same. The former depends on x *only* if X is causal for Y and corresponds to what we observe in a randomised study. The latter also depends on x, for instance when there is confounding or reverse causation, and this is what we observe in an observational study. As a simple example, let X be a binary variable indicating whether an individual's fingers are stained yellow or not, and let Y be a binary outcome for lung cancer. Since we know that stained fingers are due to smoking and smoking causes lung cancer, $p(y|x)$ describes how someone's risk of lung cancer can be predicted from inspection of their fingers. However, if we could intervene on X by staining or removing the stain from everyone's fingers, for example, $p(y|\text{do}(x))$ would no longer depend on x since finger stain in its own right does not affect lung cancer risk.

A *causal effect* is some contrast of two different interventions on X (x_1 and x_2) on the outcome Y. For continuous outcomes, the *average causal effect* (ACE), describing the average change in Y induced by setting X to be some value x_2 compared with a baseline value x_1, is an obvious causal effect parameter to consider and is the parameter that we focus on for illustrative purposes. It is defined as

$$\text{ACE}(x_1, x_2) = E(Y \mid \text{do}(X = x_2)) - E(Y \mid \text{do}(X = x_1)). \qquad (1)$$

When Y is binary, the *causal relative risk* (CRR), given by

$$\text{CRR} = \frac{P(Y = 1 \mid \text{do}(X = x_2))}{P(Y = 1 \mid \text{do}(X = x_1))},$$

or the *causal odds ratio* (COR), defined analogously, are both relevant parameters. A causal effect is identifiable if we can show mathematically, under the model assumptions and given the observable data, that the expression of the ACE in Eq. 1 – or equivalent expression for other parameters – is equal to an expression without the "do()" notation that depends purely on observational terms. Sometimes, this can be achieved by adjusting for a sufficient set of observed confounders in the usual way (18, 22). IV methods provide an alternative approach when unobserved confounding is present.

The ACE, CRR, and COR are all *population* parameters in that they are defined in terms of changes across the whole population of interest. There are other *local* causal effect parameters defined on specific subgroups of the population that we might wish to target, depending on the focus of the analysis. One well-known local effect is the "effect of the exposure on the exposed" or, the "effect

of treatment received" as it is sometimes described in a clinical trials context. Different causal parameters are identifiable under subtly different assumptions which in turn need to be justified in any given case. The interpretation of the various parameters in epidemiological applications is also open to some debate. We do not go into details here but some discussion of these issues can be found in (15, 25–27).

2.2. Instrumental Variables

Let the variable X represent the modifiable phenotype or exposure of interest, and let Y be the outcome or disease indicator, as before. We let G denote the known genetic variant associated with X which plays the role of instrument in our approach. We are interested in the causal effect of X on Y when we believe that unobserved confounding is present. Denote the unobserved confounder(s) by U. Causal inference using IV methods falls into two main categories. In order to *test* that an association is causal, we need to make certain (in)dependence assumptions concerning the four variables above. In addition, we have to make some structural assumption describing how any proposed intervention affects their joint distribution. For *estimating* the causal effect, should it seem likely that one is present, we need further parametric assumptions.

There are some *core conditions* that must be satisfied in order for the genetic variant, G, to qualify as an instrument (14, 16, 20). Using the notation $A \perp\!\!\!\perp B|C$ to mean "A is independent of B given C", these can be stated as follows:

1. $G \perp\!\!\!\perp U$ – the genetic variant is unrelated to the confounding between X and Y.
2. $G \not\perp\!\!\!\perp X$ – the genetic variant is associated with the exposure and the stronger this association, the better.
3. $G \perp\!\!\!\perp Y|(X, U)$ – given the exposure status and the confounders (if the confounders were observable), the genetic variant does not provide any additional information for the outcome, i.e. there is "no direct effect" of G on Y and no other indirect effect other than through X.

These three conditional independence assumptions define a unique directed acyclic graph (DAG) connecting the variables G, X, Y, and U as shown in Fig. 1. An equivalent statement is that

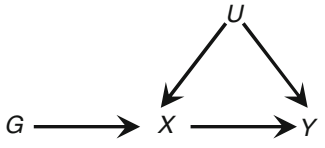

Fig. 1. The unique DAG connecting *G*, *X*, *Y*, and *U* described by the core IV conditions.

the joint distribution of the four variables factorises in the following way:

$$p(y,u,g,x) = p(y \mid x,u)p(x \mid u,g)p(u)p(g).$$

Note that these assumptions do *not* imply that $G \perp\!\!\!\perp Y|X$ or $G \perp\!\!\!\perp Y$, as has been sometimes misunderstood. Furthermore, assumptions 1 and 3 cannot be easily tested from data as they depend on U, which is unobserved, and hence have to be justified from background knowledge. In our case, the first assumption means that you must be reasonably satisfied that G is not associated with the sort of confounding you might typically expect for any particular X–Y relationship. However, Mendelian randomisation is based on the idea that genes are randomly assigned at meiosis and this implies that, across the population, genetic effects are relatively robust although not completely immune to confounding (28). Assumption 3 demands a comprehensive understanding of the underlying biological and clinical science and may be appropriately considered in a sensitivity analysis of alternative pathways.

So far, we have made assumptions about how our four variables are related "naturally". The additional structural assumption concerns what happens to the joint distribution when we intervene on X, and demands that the distributions $p(y|x, u)$, $p(g)$, and $p(u)$ are not changed by the particular intervention in X, i.e. are not changed when conditioning on $\mathrm{do}(X=x)$. This implies that the joint distribution under intervention is given by

$$p(y,u,g,x \mid \mathrm{do}(X = x^*)) = p(y \mid x^*,u)\mathbf{1}\{x = x^*\}p(u)p(g),$$

where $\mathbf{1}\{x=x^*\}$ is the indicator function taking the value 1 if $x=x^*$ and 0 otherwise. The plausibility of this assumption depends, of course, on the type of intervention being considered and needs to be justified based on background knowledge. For instance, a drug that adjusts homocysteine level might plausibly be judged to leave an individual's lifestyle behaviour unchanged. There could, however, be a placebo effect that changes the distribution of Y more than is warranted by the new value of X (homocysteine level), or the drug could affect other relevant biological processes in the body. On the other hand, if people are prevented from drinking alcohol by some change in the law, then their other health and lifestyle behaviours might change in order to compensate. Graphically, as shown in Fig. 2, intervening on X removes all

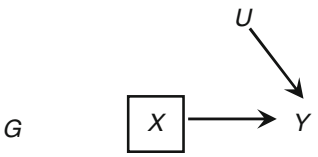

Fig. 2. The DAG representing the core IV conditions under intervention in *X*.

directed edges into X. In particular, we can see from the graph that, under intervention, we get what is often called the *exclusion restriction* (15): $G \perp\!\!\!\perp Y | \mathrm{do}(X = x)$.

Note that by deleting either the edge $U \rightarrow Y$ or $U \rightarrow X$ from Fig. 1, a test for independence of Y and G given X ($Y \perp\!\!\!\perp G|X$) is tantamount to a test for no confounding between X and Y, but we are not aware of this having been used in practice.

2.3. Testing for a Causal Effect

The three core IV conditions in Subheading 2.2, together with the structural assumption, are sufficient to test for a causal effect of X on Y *regardless* of the distributional form of the factors of the joint distribution. We do require that the joint probability distribution is *faithful* to the relevant DAG in that there are no conditional independence relationships, perhaps due to interactions that cannot be read from the graphs. Consequently, any appropriate statistical test of association between the instrument G and outcome Y amounts to a test for a causal effect of X on Y. (See ref. 16 for a detailed discussion.)

2.3.1. Homocysteine and Stroke

Is there a causal relationship between high plasma homocysteine concentrations and risk of stroke (29)? The T allele of the MTHFR C677T polymorphism is known to be associated with homocysteine levels with TT homozygotes having higher levels than CC homozygotes, in particular. A summary estimate of the association between homocysteine levels and risk of stroke (X–Y relationship) from a meta-analysis gave an odds ratio of 1.59 corresponding to a 5 μmol/L observed increase in homocysteine. Dichotomising MTHFR into TT and CC carriers, the odds ratio for the genotype–stroke association (G–Y) was 1.26 and found to be significant. The conclusion is that a causal effect of homocysteine level on risk of stroke is plausible. However, *none* of the reported values give any indication of the *size* of this effect as the homocysteine–stroke association could be confounded.

2.3.2. Plasma Fibrinogen and CHD

Do higher plasma fibrinogen levels increase the risk of coronary heart disease (CHD) (30)? Various observational studies have reported increased risk associated with higher fibrinogen levels for various cardiovascular outcomes. The G-455$\rightarrow A$ polymorphism in the promoter region of the β-fibrinogen gene is consistently associated with differences in fibrinogen levels and plays the role of the instrument. A meta-analysis of 16 studies produced a "per allele" odds ratio of 0.96 with associated 95% confidence interval of (0.89, 1.04). The conclusion is that there is no support for a causal effect of fibrinogen levels on CHD; or in other words, if fibrinogen has a causal effect then it is too small to be detected in this meta-analysis of 16 studies.

2.3.3. Alcohol Consumption and Blood pressure

The previous examples concerned the effect of some intermediate phenotype on a disease. We can also use the idea of Mendelian randomisation when we have a modifiable exposure, such as alcohol consumption, for which positive (e.g. CHD) and negative (e.g. liver cirrhosis and some cancers) effects have been reported in observational studies. Besides being difficult to measure due to reporting bias, alcohol consumption is strongly associated with all kinds of confounding factors, and so there are doubts about the causal nature of any of the above associations (31). We consider the issue of whether there is a causal effect of alcohol consumption on blood pressure.

The ALDH2 gene determines blood acetaldehyde, the principal metabolite for alcohol, and is known to be associated with alcohol consumption. In particular, individuals homozygous for the "null" variant *2 suffer unpleasant symptoms, such as facial flushing, nausea, drowsiness, and headache after alcohol consumption. Heterozygotes have a limited ability to metabolise acetaldehyde but have a less severe reaction than *2*2 homozygotes. Consequently, *2*2 homozygotes have lower alcohol consumption than the "wild type" *1*1 homozygotes *regardless* of their other lifestyle behaviours while heterozygotes tend to drink intermediate amounts. In the meta-analysis of Chen et al. (31), there was no apparent association between ALDH2 and typical confounding factors that one would expect for the alcohol–blood pressure relationship. This, together with the random allocation of genes at conception makes us fairly confident about core IV assumption 1. Current knowledge of the biochemical function of ALDH2 excludes the possibility that it could be associated with blood pressure via another pathway besides alcohol consumption (core IV assumption 3).

Blood pressure was found to be 7.44 mmHg higher on average for *1*1 homozygotes than for *2*2 homozygotes with 95% CI (5.39, 9.49) yielding a p value of $p = 1.1 \times 10^{-12}$ for high versus low consumption. Blood pressure was 4.24 mmHg higher on average for *1*2 heterozygotes than for *2*2 homozygotes with 95% CI (2.18, 6.31) giving a p value of $p = 0.00005$ for moderate versus low consumption. Most of the studies were on Japanese populations (where ALDH2*2*2 is common) so these results are for males as Japanese women drink very little alcohol in general. The fact that there was no observed relationship between genotype and blood pressure for women indicated that the above association is indeed due to alcohol consumption, for which ALDH2 is a proxy, and not due to the gene itself or some alternative pathway by which ALDH2 might predict blood pressure. The highly significant association between the ALDH2 variant and blood pressure is strong evidence of a causal effect. In fact, contrary to reported observational claims, it would appear that even moderate drinking can be harmful.

2.4. Estimating a Causal Effect

Once a test indicates that a causal effect is likely, we would typically want to know the size of this effect. This is more difficult. When all variables are binary, or categorical, only upper and lower *bounds* on the causal effect can be calculated without any extra assumptions (32). The width – and hence usefulness – of these bounds depend on the strength of the IV and the amount of confounding, but they do give an idea of how informative the data are. Hence, the IV core conditions and structural assumption are not sufficient for point identification of causal parameters and extra parametric assumptions are required.

When the ACE of Eq. 1 is of interest and Y is continuous (possibly suitably transformed), it is popular to assume *linearity* of all relationships and *no interactions*. The structural (causal) assumption only appears in the regression of Y on X and U:

$$E(Y \mid X = x, U = u) = E(Y \mid \mathrm{do}(X = x), U = u) = \mu + \beta x + \delta u$$

This yields $\beta = \mathrm{ACE}(x+1, x)$ as the relevant causal parameter. Note that the above assumes that there is no effect modification of the effect of X on Y by U on the linear scale, i.e. people in various subpopulations, like men/women or older/younger people, all react in the same way to exposure. The parameter β cannot be estimated from the above regression as U is unobserved. Likewise, we cannot ignore U and estimate it from a regression of Y on X as this would give a biased estimate due to the collinearity $U \not\perp\!\!\!\perp X$. From the regression of X on G and U

$$E(X \mid G = g, U = u) = \eta + \alpha g + \zeta u,$$

we *can* estimate α by ignoring U since $G \perp\!\!\!\perp U$. It is easy to show (16) that

$$E(Y \mid G = g) = \tilde{\mu} + \alpha\beta \cdot g,$$

so $\alpha\beta$ can be estimated from a regression of Y on G. Hence, a consistent estimator for $ACE(x+1, x) = \beta$ is given by the ratio of the estimated coefficients, $\hat{\beta}_{Y|G}$ and $\hat{\beta}_{X|G}$ from the regressions of Y on G and of X on G, respectively. It is useful that these could even be estimated from separate studies, one where X, G are observed and another one where Y, G are observed. In this situation, the above ratio estimator is equivalent to the popular "two-stage least squares" (2SLS) estimator which regresses Y on values of X predicted from the "first-stage" regression of X on G and the terms are often used interchangeably.

In an investigation into the causal effect of circulating C-reactive protein (CRP) and the metabolic syndrome, three-SNP haplotypes from the CRP gene were used as instruments for associations between serum CRP levels and various metabolic syndrome phenotypes (33). For one particular outcome – insulin resistance measured by homoeostasis model assessment (HOMA-R) – a clear

observational association was reported with a doubling of CRP levels leading to a significant increase of about 8% in HOMA-R ($p < 0.0001$). But CRP is known to be associated with a wide range of lifestyle and socioeconomic characteristics. Moreover, it could be elevated as a result of atherosclerosis or insulin resistance, so confounding and reverse causation cannot be excluded. The core IV assumptions appear to be reasonably satisfied, although not enough is known about the biological pathways involving CRP to be fully sure about assumption 3. Standard checks for linearity on log(CRP) and log(HOMA-R) looked reasonable. It is, of course, impossible to check any parametric assumptions about the unobserved confounders, especially the one of no effect modification. This has to be justified with subject matter background knowledge, instead. The 2SLS approach, using the regressions of log(HOMA-R) on the CRP haplotypes and of log(CRP) on the CRP haplotypes, estimated that doubling CRP levels *reduces* the HOMA-R score by 6% ($p > 0.1$). Since the result is non-significant, we conclude that the data do not support a causal effect of circulating levels of CRP on insulin resistance. This result appears to contradict the naive analysis, which may indeed be due to confounding and reverse causation.

3. Further Issues and Complications

There are well-known problems with the 2SLS-estimator. The standard deviation of the estimator is typically much larger than that of the estimator obtained from a naive regression of Y on X. This is especially so when we have a weak instrument, i.e. when $Corr(G, X) \approx 0$ so that IV is not very informative for X. Note that it is impossible to find a strong instrument when there is a lot of confounding. One notable problem is that the assumption of linearity cannot be true when Y is binary, although it could be a good approximation over a particular range of exposure levels in some cases. This is an issue for epidemiological applications since many outcomes of interest are naturally binary.

The main problem for the non-linear case is that the relationship between the two regressions (Y on G, and X on G) and the relevant causal parameter, i.e. CRR or COR, is no longer straightforward and any estimators derived from these are biased (16, 27). There are other IV methods that can yield estimates of certain causal effects for binary outcomes, but they all require strong additional assumptions (15, 34–36). It is important to note that different approaches target different causal parameters in the sense that they estimate individual, local, or population effects. Some estimators, such as those derived from structural mean modelling approaches, also require joint observation of all three variables (G, X, and Y) for all individuals, whereas the "Wald-type" estimators based on ratios of differences (of which

2SLS is an example), do not (27). This has implications for meta-analyses as not all studies typically supply joint observations. Structural mean models make weaker assumptions than the other approaches in that a parametric model for the regression of X on G and U is not required. However, these approaches target the local effect of exposure on the exposed and are only unbiased for a population effect with additional assumptions.

Violations of the core IV conditions are also possible and these have implications even for testing for a causal relationship. The most important and likely violation occurs when there is population stratification, where we have different allele frequencies in subpopulations which may in turn also differ in their lifestyles (giving rise to an association between G and U) or in their disease risk (giving rise to an association between G and Y not screened off by X, U). A sensible study design should take this possibility into account. The chosen instrument could also be in linkage disequilibrium (LD) with another variant which is associated with the disease via a route other than through its effect on X, the exposure of interest. Likewise, problems can be caused by pleiotropic effects and canalisation or developmental compensation (1, 10, 37). If only insufficient prior knowledge about the genetic or confounding mechanisms is available to justify the core conditions, results that seem to indicate a causal effect may very well have an alternative, non-causal explanation that we are not aware of. DAGs can be used to represent what is believed about the biology and then be queried regarding the validity of our assumptions (12, 16). For example, the genetic variant chosen as instrument may not be *the* causal gene for the exposure of interest but is in LD with a causal gene which is unobserved. This could be thought of as *measurement error* in the genetic data. However, as illustrated in Fig. 3, this does not necessarily imply any violations of the core IV conditions. G_1 might not be as good an instrument as G_2 in the sense that its association with X is weaker, but it is (1) independent of U, (2) associated with X, and (3) conditionally independent of Y given X and U.

Finding a genetic variant that is a suitable IV is also problematic, and there are currently not very many well-studied variants for the typical exposures of interest in epidemiology. Genetic variants that arise from genome-wide association studies could be problematic in that the gene–phenotype associations are often weak

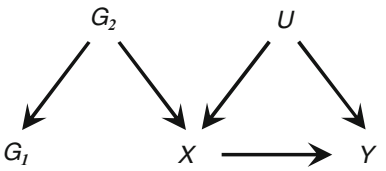

Fig. 3. The chosen instrument G_1 is not causal for X but is associated with another genetic variant, G_2, which is driving all the association. All IV core conditions are satisfied for G_1.

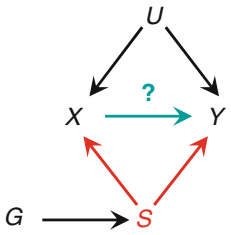

Fig. 4. $G \not\perp\!\!\!\perp X$ and $G \perp\!\!\!\perp U$ but the IV core conditions are not satisfied.

and may not even be reproducible. Even when strong, reproducible associations are found, we then have to be convinced that enough is known about the functionality of the gene in order to claim that the core conditions are satisfied for an IV analysis. Such knowledge does not derive from an association study. On the positive side, thanks to the current rapid advances in functional genomics, the required information on such variants is gradually being accrued. Figure 4 depicts a situation, where we have a clear association between G and X, we can argue the independence between G and U but without understanding the functionality of the gene, there is no way of knowing that the third condition is violated. Since an association between G and Y is evident, we would incorrectly deduce that X causally affects Y, whereas an alternative explanation for this association is that the gene causes an unobservable health problem S which then affects both X and Y.

4. Conclusion

A Mendelian randomisation analysis is *not* aimed at identifying genetic factors that are causal for disease risk. On the contrary, the method requires a known and well-understood genetic variant in order to facilitate causal inference about the effect of an exposure on the disease of interest. One of the limitations for IV methods is finding valid instruments. This is also an issue with genetic instruments in our applications but is hopefully becoming less so with the recent rapid advances in genetic epidemiology (8). Inferring causality from observational data is problematic, but we would argue that some of the confusion about misleading results from observational studies stems from the lack of clear delineation between the notions of association and causation, at a conceptual as well as formal level (24). Only when this distinction is made explicit, can we identify and understand the crucial assumptions that permit a causal interpretation. Only then are we able to critically scrutinise these assumptions, justify or reject them, and hence assess the practical impact of any results. Solid

background knowledge is essential for causal analysis. With Mendelian randomisation, we have an advantage over many other areas of application of IV methods in that genetics provide a rich source of information.

Acknowledgement

We acknowledge research support for all authors from the Medical Research Council through a collaborative project grant (G0601625) addressing causal inference in observational epidemiology using Mendelian randomisation.

References

1. D.A. Lawlor, R.M. Harbord, J.A.C. Sterne, N. Timpson, and G.D. Smith. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, 27:1133–1328, 2008.
2. R. Doll, R. Peto, J. Boreham, and I. Sutherland. Mortality from cancer in relation to smoking: 50 years observations on British doctors. *British Journal of Cancer*, 92:426–429, 2005.
3. MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the Medical Research Council vitamin study. *Lancet*, 338:131–137, 1991.
4. T.O. Scholl and W.G. Johnson. Folic acid: influence on the outcome of pregnancy. *American Journal of Clinical Nutrition*, 71 (Suppl.):12955–13035, 2000.
5. W.C. Willett. Vitamin A and lung cancer. *Nutritional Review*, 48:201–211, 1990.
6. R. Peto, R. Doll, J.D. Buckley, and M.B. Sporn. Can dietary beta-carotene materially reduce human cancer rates? *Nature*, 290:201–208, 1981.
7. Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New England Journal of Medicine*, 330:1029–1035, 1994.
8. G.D. Smith, S. Ebrahim, S. Lewis, A.L. Hansell, L.J. Palmer, and P.R. Burton. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet*, 366:1484–1498, 2005.
9. A. Tatsioni, N.G. Bonitis, and J.P.A. Ioannidis. Persistence of contradicted claims in the literature. *Journal of the American Medical Association*, 298:2517–2526, 2007.
10. G.D. Smith and S. Ebrahim. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, 32:1–22, 2003.
11. M.B. Katan. Commentary: Mendelian randomization, 18 years on. *International Journal of Epidemiology*, 33:10–11, 2004.
12. N.A. Sheehan, V. Didelez, P.R. Burton, and M.D. Tobin. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Medicine*, 5:e177, 2008.
13. R.J. Bowden and D.A. Turkington. *Instrumental Variables*. Cambridge University Press, Cambridge, 1984.
14. S. Greenland. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, 29:722–729, 2000.
15. M.A. Hernán and J.M. Robins. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*, 17:360–372, 2006.
16. V. Didelez and N.A. Sheehan. Mendelian randomisation as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, 16:309–330, 2007.
17. J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82:669–710, 1995.
18. S.L. Lauritzen. Causal inference from graphical models. In O.E. Barndorff-Nielsen, D.R. Cox, and C. Kluppelberg, editors, *Complex Stochastic Systems*, Chapter 2, 63–107. Chapman & Hall, Boca Raton, 2000.
19. A.P. Dawid. Influence diagrams for causal modelling and inference. *International Statistical Review*, 70:161–189, 2002.

20. A.P. Dawid. Causal inference using influence diagrams: the problem of partial compliance. In P.J. Green, N.L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*, 45–81. Oxford University Press, Oxford, 2003.
21. D.B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.
22. J. Pearl. *Causality*. Cambridge University Press, Cambridge, 2000.
23. M.A. Hernán. A definition of causal effect for epidemiologic research. *Journal of Epidemiology and Community Health*, 58:265–271, 2004.
24. V. Didelez and N.A. Sheehan. Mendelian randomisation: why epidemiology needs a formal language for causality. In F. Russo and J. Williamson, editors, *Causality and Probability in the Sciences*, volume 5, *Texts in Philosophy*, 263–292. London College Publications, London, 2007.
25. J.M. Robins, T.J. VanderWeele, and T.S. Richardson. Comment on: Causal effects in the presence of non compliance: a latent variable interpretation. *Metron*, 64:288–298, 2006.
26. S. Geneletti and A.P. Dawid. The effect of treatment on the treated: a decision theoretic perspective. In P. McKay Illari, F. Russo and J. Williamson, editors, *Causality in the Sciences*,Oxford University Press, 2010.
27. V. Didelez, S. Meng, and N.A. Sheehan. Assumptions of IV methods for observational epidemiology. Statistical Science, 25: 22-40, 2010.
28. G.D. Smith, D.A. Lawlor, R. Harbord, N. Timpson, I. Day, and S. Ebrahim. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Medicine*, 4:e352, 2007.
29. J.P. Casas, L.E. Bautista, L. Smeeth, P. Sharma, and A.D. Hingorani. Homocysteine and stroke: evidence on a causal link from Mendelian randomisation. *Lancet*, 365:224–232, 2005.
30. G.D. Smith, R. Harbord, J. Milton, S. Ebrahim, and J. Sterne. Does elevated plasma fibrinogen increase the risk of coronary heart disease? *Arteriosclerosis, Thrombosis and Vascular Biology*, 25:2228–2233, 2005.
31. L. Chen, G.D. Smith, R. Harbord, and S.J. Lewis. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Medicine*, 5:e52, 2008.
32. A.A. Balke and J. Pearl. Counterfactual probabilities: computational methods, bounds and applications. In R.L. Mantaras and D. Poole, editors, *Proceedings of the 10th Conference on Uncertainty in Artificial Inteligence*, 46–54, 1994.
33. N.J. Timpson, D.A. Lawlor, R.M. Harbord, T.R. Gaunt, I.N.M. Day, L.J. Palmer, A.T. Hattersley, S. Ebrahim, G.D.O. Lowe, A. Rumpley, and G.D. Smith. C-reactive protein and its role in metabolic syndrome: a Mendelian randomisation study. *Lancet*, 366:1954–1959, 2005.
34. J.M. Robins. Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics*, 23:2379–2412, 1994.
35. S. Vansteelandt and E. Goetghebeur. Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society, Series B*, 65:817–835, 2003.
36. P. Clarke and F. Windmeijer. Instrumental variable estimators for binary outcomes. Working Paper 10/239, Centre for Market and Public Organisation, University of Bristol, 2010.
37. D. Nitsch, M. Molokhia, L. Smeeth, B.L. DeStavola, J.C. Whittaker, and D.A. Leon. Limits to causal inference based on Mendelian randomization: a comparison with randomised controlled trials. *American Journal of Epidemiology*, 163:397–403, 2006.