Statistical Methods in Medical Research 2007; **16**: 309–330

Mendelian randomization as an instrumental variable approach to causal inference

Vanessa Didelez Departments of Statistical Science, University College London, UK and
Nuala Sheehan Departments of Health Sciences and Genetics, University of Leicester, UK

In epidemiological research, the causal effect of a modifiable phenotype or exposure on a disease is often of public health interest. Randomized controlled trials to investigate this effect are not always possible and inferences based on observational data can be confounded. However, if we know of a gene closely linked to the phenotype without direct effect on the disease, it can often be reasonably assumed that the gene is not itself associated with any confounding factors – a phenomenon called *Mendelian randomization*. These properties define an instrumental variable and allow estimation of the causal effect, despite the confounding, under certain model restrictions. In this paper, we present a formal framework for causal inference based on Mendelian randomization and suggest using directed acyclic graphs to check model assumptions by visual inspection. This framework allows us to address limitations of the Mendelian randomization technique that have often been overlooked in the medical literature.

1 Introduction

Inferring causation from observed associations is often a problem with epidemiological data as it is not always clear which of two variables is the cause, which the effect, or whether both are common effects of a third unobserved variable. In the case of experimental data, causal inference is facilitated either by using randomization or experimental control. By randomly allocating levels of 'treatment' to 'experimental units', the randomized experiment of Fisher[1] renders reverse causation and confounding highly unlikely. In a controlled experiment, causality can be inferred by the experimental setting of all other variables to constant values although Fisher[2] argued that this is inferior to randomization as it is logically impossible to know that 'all other variables' have been accounted for. In many biological settings, it is not possible to randomly assign values of a hypothesized 'cause' to experimental units for ethical, financial or practical reasons. In epidemiological applications, for example, randomized controlled trials (RCTs) to evaluate the effects of exposures such as smoking, alcohol consumption, physical activity and complex nutritional regimes are unlikely to be carried out. However, when randomization *is* possible, the number of spurious causal associations reported from conventional observational epidemiological studies that have failed to be replicated in

Address for correspondence: Nuala Sheehan, Department of Health Sciences and Genetics, University of Leicester, Leics LE1 6TP, UK. E-mail: nas11@le.ac.uk

10.1177/0962280206077743

large-scale follow-up RCTs, such as the association between beta-carotere and smoking-related cancers[3,4], for example, is concerning. One of the main reasons for such spurious findings is *confounding* whereby an exposure is associated with a range of other factors affecting disease risk, like socioeconomic position or behavioural choices. Controlling for confounding is problematic as one can never really know what the relevant confounders are (cf. Fisher's argument above). Furthermore, accurate measurements of typical confounders can be difficult to obtain.

Mendelian randomization has been proposed as a method to test for, or estimate, the causal effect of an exposure or phenotype on a disease when confounding is believed to be likely and not fully understood.[5,6] It exploits the idea that the genotype only affects the disease status indirectly and is assigned randomly (given the parents' genes) at meiosis, independently of the possible confounding factors. It is well known in the econometrics and causal literature,[7–9] and slowly being recognized in the epidemiological literature[10–12] that these properties define an *instrumental* variable (IV). Our claim here is that they are *minimal* conditions in the sense that unique identification of the causal effect of the phenotype on the disease status is only possible in the presence of additional fairly strong assumptions. This has often been overlooked in the medical literature. Additional assumptions can take the form of *linearity* and *additivity* assumptions, as are typically assumed in econometrics applications, or assumptions about the *compliance* behaviour of subjects under study, as are often made in the context of randomized trials with incomplete compliance.[8] Without such assumptions it is only possible to compute bounds on the causal effect,[13–15] typically when all relevant variables are binary.

We begin with a brief description of Mendelian randomization before introducing the causal concepts required to establish the role of IVs. We will then show how an IV can be used to test for and estimate the causal effect and conclude with a discussion of problems and open questions.

2 Mendelian randomization

The term Mendelian randomization, as we use it here, derives from an idea put forth by Katan.[16] In the mid-1980s, there was much debate over the direction of an association between low serum cholesterol levels and cancer, both in observational studies and in the early trials on lowering cholesterol. The hypothesis was that low serum cholesterol increases risk of cancer but it was also possible that either the presence of hidden tumours induces a lowering of cholesterol in future cancer patients or other factors such as diet and smoking affect both cholesterol levels and cancer risk. The observation that people with the rare genetic disease abetalipoproteinaemia, resulting in extremely low serum cholesterol levels, do not display a tendency to cause premature cancer led to the idea that identification of a larger group of individuals genetically predisposed to having low cholesterol levels might help to resolve the issue. The apolipoprotein E (APOE) gene was known to be associated with serum cholesterol levels. The E2 allele is associated with lower levels than the other two alleles, E3 and E4, so E2 carriers should have relatively low levels of serum cholesterol and, crucially, should be similar to E3 and E4 carriers in socioeconomic position, lifestyle and all other respects. Since lower cholesterol levels in

E2 carriers are present from birth, Katan reasoned that a prospective study is unnecessary and a simple comparison of APOE genotypes in cancer patients and controls should suffice to resolve the causal dilemma. If low serum cholesterol level is really a risk factor for cancer, then patients should have more E2 alleles and controls should have more E3 and E4 alleles. Otherwise, APOE alleles should be equally distributed across both groups.

In short, the idea is to test the hypothesis of a causal relation between serum cholesterol levels and cancer by studying the relationship between cancer and a genetic determinant of serum cholesterol. The former association is affected by confounding, the latter is not since alleles are assigned at random conditionally on the parents' alleles. Causality can be inferred because we are more or less back in the world of Fisher's randomized experiment although the analogy with RCTs is much more approximate in population-based genetic association studies than it is in parent-offspring designs, for instance.[5] Unlike genetic epidemiology, the aim in a Mendelian randomization analysis is not to identify groups of individuals at risk on the basis of their genotype but to study the genotype because it mimics the effect of some exposure of interest. While Katan's original idea was centred around hypothesis testing to confirm or disprove causality, the method is now also applied to estimate the size of the effect of the phenotype on the disease together with a measure of its uncertainty[17] and, indeed, to compare this estimate with that obtained from observational studies in order to assess the extent to which the observational studies have controlled for confounding. Katan's idea was never actually implemented but the subsequent statin trials on the effects of *high*-cholesterol levels on coronary heart disease (CHD) risk, disproved the original hypothesis.[18,19]

3 Causal concepts and terminology

Accounts of Mendelian randomization feature frequent use of causal vocabulary to express something that is more than association between genotypes, intermediate phenotypes and disease. While this is common practice in the medical literature where underlying knowledge about the biology of the problem may indeed allow one to deduce the direction of an observed association and where 'causal pathways' for disease are familiar concepts, it is important for our purposes that we make a formal distinction between association and causation.

3.1 Interventions

As elsewhere in the literature,[20–22] we regard causal inference to be about predicting the effect of *interventions* in a given system. For the applications we are considering, this would typically constitute a public health intervention such as adding folate to flour, vitamin E to milk or giving advice on diet, and so on. There are many other notions of causality including, for instance, its use in a courtroom for retrospective assignment of guilt, but we will not consider those here.

Let X be the cause under investigation, for example, cholesterol or homocysteine level, and Y the response, that is, the disease status, such as cancer or coronary heart disease. By intervening on X, we mean that we can set X (or more generally its distribution)

to any value we choose without affecting the distributions of the other variables in the system, other than through the resulting changes in X. This is clearly an idealistic situation and not always justified for the examples of public health interventions given above. For example, increasing dietary folate will not determine a specific homocysteine level (see Davey Smith and Ebrahim[5] for a discussion of this example) which is why we need results from a controlled randomized trial on the effect of adding folate to the diet to inform the intervention. However, a causal analysis exploiting Mendelian randomization can be used to generate hypotheses that can afterwards be investigated by controlled randomized trials where applicable. Moreover, if a phenotype is found to be causal in the above sense, different ways of intervening on this phenotype can then be explored.

3.2 Causal effect

The causal effect is a function of the distributions of Y under different interventions in X denoted as $p(y|do(X = x))$ by Pearl.[9] It is well known that this is not necessarily equal to the usual conditional distribution $p(y|X = x)$ which just describes a statistical dependence.[9,21] The different notations are reflecting the common phrase 'correlation is not causation'. (In the Appendix we show, for example, that $E(Y|do(X = x))$ is not the same as $E(Y|X = x)$ in a linear model with confounding.)

We define the (average) causal effect (ACE) as the difference in expectations under different settings of X:

$$\text{ACE}(x_1, x_2) = E(Y|do(X = x_1)) - E(Y|do(X = x_2)) \quad (1)$$

where x_2 is typically some baseline value. In particular, X is regarded as causal for Y if the ACE (Equation (1)) is non-zero for some values x_1, x_2. If X is binary, the unique ACE is given by $E(Y|do(X = 1)) - E(Y|do(X = 0))$. If Y is continuous, a popular assumption is that the causal dependence of Y on X is linear (possibly after suitable transformations), that is $E(Y|do(X = x)) = \alpha + \beta x$. In this case, the ACE is $\beta(x_1 - x_2)$ and can be simply summarized by β. (See Section 6.1 for more discussion on the linear case.) In the cases of more than two categories and/or non-linear dependence, the ACE is not necessarily summarized by a single parameter and one may want to choose a different causal parameter altogether (Section 6.3). In many applications it makes sense to also condition on covariates in Equation (1) in order to investigate the causal effect within specific subgroups, for example, age groups or people with specific medical histories.

3.3 Identifiability

A causal parameter is identifiable if it is unique given the distribution of the observable variables. Mathematically, this amounts to being able to express the parameter in terms that do not involve the intervention (i.e., the 'do' operation) by using 'observational' terms only. As noted above, $p(y|do(X = x))$ is not necessarily the same as $p(y|X = x)$ due to confounding, for example. Hence we cannot easily estimate parameters of $p(y|do(X = x))$ from observations that represent $p(y|X = x)$. In the rare case of known confounders, it can be shown that the intervention distribution can be re-expressed in observational terms and can hence be estimated from the observed data by adjusting

for these confounders; graphical rules can be used to determine the variables to adjust for.[20–22] The IV technique based on Mendelian randomization allows a different way of identifying causal parameters when the confounders are unobservable.

4 Instrumental variables

We now define the core conditions that characterize an instrumental variable (IV). These properties have been given in many different forms. Our terminology and notation closely follow Greenland[10] and Dawid.[23] Other authors use counterfactual variables[8,13] or linear structural equations.[9,24] We express these properties as conditional independence statements where $A \perp\!\!\!\perp B|C$ means 'A is conditionally independent of B given C'. On their own, they do not allow identification of the ACE as we will discuss more fully later. For now, we just present these conditions together with a graphical way of depicting and checking the relevant conditional independencies.

4.1 Core conditions

Let X and Y be defined as above with the causal effect of X on Y being of primary interest. Furthermore, let G be the variable that we want to use as the instrument (the genotype in our case) and let U be an unobservable variable that will represent the confounding between X and Y. The 'core conditions' that G has to satisfy are the following:

1) $G \perp\!\!\!\perp U$, that is, G must be (marginally) independent of the confounding between X and Y;
2) $G \not\perp\!\!\!\perp X$, that is, G must not be (marginally) independent of X;
3) $Y \perp\!\!\!\perp G \mid (X, U)$, that is, conditionally on X and the confounder U, the instrument and the response are independent.

These assumptions cannot easily be tested and have to be justified on the basis of subject matter or background knowledge. This is because U, by definition, is not observable: if it were, we could adjust for it and would not need any instrument to identify the effect of X on Y. Furthermore, the above assumptions do not imply testable conditional independencies regarding the instrument G. In particular they do *not* imply that G is independent of Y, either marginally or conditionally on X alone.

4.2 Graphical representation

Graphical models based on directed acyclic graphs (DAGs) can be used to represent conditional independencies amongst a set of joint variables in the following way. Every node of the graph represents a variable and these can be linked by directed edges which we represent as arrows ($\longrightarrow$). If $a \longrightarrow b$ we say that a is a *parent* of b and b is a *child* of a. If $a \longrightarrow \cdots \longrightarrow b$ then a is an *ancestor* of b and b is a *descendant* of a. A cycle occurs when there exists an unbroken sequence of directed edges leading from a back to itself. DAGs have no such cycles. All the conditional independencies represented in the graph can be derived from the *Markov properties* of the graph by which every node is independent of all its non-descendants given its parents.[25,26]

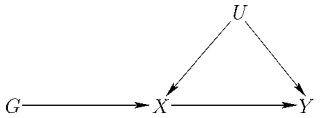

Figure 1 DAG representing the core conditions for an instrument.

Figure 1 shows the unique DAG involving G, X, Y and U that satisfies the core conditions 1)–3) of Section 4.1. From the graph, we have that $G \perp\!\!\!\perp U$ because G and U are non-descendants of each other (and their parent sets are empty). Likewise, $G \not\perp\!\!\!\perp X$ because X is a descendant of G, and $Y \perp\!\!\!\perp G|(X, U)$ because (X, U) are the parents of Y and G is a non-descendant of Y. The conditional independence restrictions imposed by the graph in Figure 1 are equivalent to a factorization of the joint density in the following way:

$$p(y, x, u, g) = p(y|u, x)p(x|u, g)p(u)p(g) \tag{2}$$

From this it can be seen (by integrating out y and conditioning on x) that $G \not\perp\!\!\!\perp U|X$, for instance. Similarly, by integrating out x and conditioning on y, we have that $G \not\perp\!\!\!\perp U|Y$, or formally

$$p(g, u|y) = \frac{p(u)p(g)\sum_x p(y|u, x)p(x|u, g)}{\sum_{u,g} p(u)p(g)\sum_x p(y|u, x)p(x|u, g)} = \frac{p(u)p(g)p(y|u, g)}{p(y)} \neq p(g|y)p(u|y)$$

even though marginally $p(g, u) = p(g)p(u)$. This is the so-called *selection effect* whereby two variables such as G and U, which are marginally independent, may become dependent once we condition on a common descendant. The selection effect is particularly relevant to case–control data when everything is conditional on the outcome Y. In graphical terms, a *moral edge* is induced between two variables that have a common child when conditioning on this child or a descendant thereof (Cowell *et al.* [26], for example). Here, as G and U have a common child X, and the variable we condition on, Y, is a descendant of X, such a moral edge is required to represent the case–control situation (Figure 2). Particular consideration will therefore be given to the suitability of Mendelian randomization for case–control data.

Note that DAGs only represent conditional dependencies and independencies: they are not causal in themselves despite the arrow suggesting a 'direction' of dependence. We say that the DAG has a *causal* interpretation with respect to the relationship between X and Y, or, more specifically, the DAG is causal with respect to X, if we believe that an intervention in X does not affect any of the other factors in the joint distribution of Equation (2)[9], that is,

$$p(y, u, g|do(X = x_0)) = p(y|u, x_0)p(u)p(g) \tag{3}$$

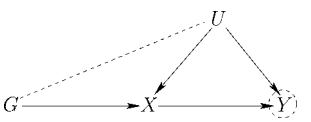

Figure 2 Conditioning on the outcome Y possibly induces an association between G and U.

Graphically, the intervention corresponds to removing all the arrows leading into X from the graph in Figure 1[9]. Note that the validity of the assumption about the intervention that allows Equation (3) depends on the variables included in the graph and the actual kind of intervention being contemplated. After all, why should the conditional distributions of the remaining variables remain unchanged, in general, if a potentially very different situation is created by intervening? For this to be plausible, the graph typically needs to be augmented by additional variables that might be relevant and U specified in more detail in order to represent what is thought to be the data generating process based on subject matter knowledge. An example of this is given in the next subsection and several more in Section 7.

4.3 *G–X* association

Core condition 2) states that G and X need to be associated. The stronger this association the 'better' G as an instrument providing more information on the causal effect of X on Y in the sense of small standard errors and narrow confidence intervals; this is well known for the linear case discussed in Section 6.1. In the extreme case where G almost fully determines X, knowing that G is independent of U means that X cannot depend much on U and thus confounding is low. Martens *et al*.[11] discuss the functional relation between strength of instrument and amount of confounding for the linear case, and show that it is impossible to find a strong instrument when the amount of confounding is high. In practice, the requirement that G and X be strongly associated may also pose a problem as we are limited by what is known from genetic studies and many phenotypes of interest may not have a strong, or well understood, genetic component.

It is important to note that for G and X to be associated, the instrument G does not need to be causal for X, that is, the arrow $G \to X$ in Figure 1 is not representing a causal relation. (See also Hernán and Robins[12].) The association could instead be due to mediation or via another unobserved variable that affects both G and X. This is illustrated in the graph of Figure 3 where we consider two genotypes, G_1 and G_2, which might be related by linkage disequilibrium (cf. also Section 7). The conditional independencies encoded by this graph are

$$(G_1, G_2) \perp\!\!\!\perp U, \quad X \perp\!\!\!\perp G_1 | (G_2, U), \quad Y \perp\!\!\!\perp (G_1, G_2) | (X, U)$$

where the first and second conditional independencies together imply $X \perp\!\!\!\perp G_1 | G_2$[27]. The corresponding factorization of the density is given by

$$p(y, x, u, g_1, g_2) = p(y|x, u)p(x|u, g_2)p(g_1|g_2)p(g_2)p(u)$$

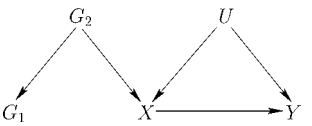

Figure 3 An alternative formulation to the IV conditions.

If we believe that we can intervene in X without affecting anything else, we have

$$p(y, u, g_1, g_2|do(X = x_0)) = p(y|x_0, u)p(g_1|g_2)p(g_2)p(u) \tag{4}$$

Now, assuming that only G_1, not G_2, is observed, the joint distribution of the remaining variables is

$$p(y, x, u, g_1) = p(y|x, u)p(x|u, g_1)p(g_1)p(u)$$

This is the same as the factorization in Equation (2) and equivalent to assuming our core conditions 1)–3) with only G_1 as IV. It also yields the same intervention distribution as before

$$p(y, u, g_1|do(X = x_0)) = p(y|x_0, u)p(g_1)p(u)$$

which is alternatively obtained by integrating out g_2 in Equation (4). For our purposes therefore, we do not have to find the 'right' gene as it does not matter how the association between X and the genotype comes about. Hence, without loss of generality, we will assume the situation depicted in Figure 1 and described algebraically in Equation (2), and in Equation (3) for the intervention case. However, as mentioned above, the stronger the $G - X$ association the better the instrument.

5 Testing for zero causal effect

Assuming that the core conditions 1)–3) of Section 4.1 are satisfied, let us first consider the situation where we just want to know whether there is a causal link from a phenotype X to a disease Y without quantifying it. In this section, we investigate whether a test for dependence between X and Y can be replaced by a test of dependence between the instrument G and Y. Because of the selection effect in case–control data mentioned above, we consider prospective and retrospective views separately.

5.1 Prospective view

From Equation (3) we obtain the distribution of Y under intervention as

$$p(y|do(X = x_0)) = \sum_{u,g} p(y|u, x_0)p(u)p(g) = \sum_{u} p(y|u, x_0)p(u) \tag{5}$$

This can be recognized as the usual adjustment formula if U were observable, that is, we partition the population according to U, assess the effect of X on Y within each subgroup and then average over the subgroups (Pearl[9], p. 78). The ACE, as defined by Equation (1), is then

$$\text{ACE}(x_1, x_2) = \sum_{u} (E(Y|U = u, X = x_1) - E(Y|U = u, X = x_2))p(u)$$

If $E(Y|U = u, X = x) = E(Y|U = u)$, or more strongly if $p(y|u, x) = p(y|u)$ (i.e., $Y \perp\!\!\!\perp X|U$) then the causal effect is obviously zero. Note that $Y \perp\!\!\!\perp X|U$ has a graphical

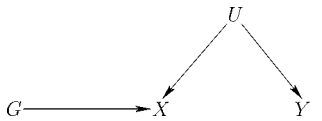

Figure 4 DAG representing $Y \perp\!\!\!\perp X \mid U$.

counterpart, shown in Figure 4, which is obtained by deleting the arrow from X to Y in Figure 1. However, the reverse is not necessarily true. If the ACE is zero, or formally if $p(y|do(X = x))$ does not depend on x, then we cannot conclude that $p(y|u, x)$ does not depend on x because, as implied in Equation (5), there could be an interaction between X and U in their effect on Y which together with the weights $p(u)$ cancels out the overall effect. This is the problem of 'non faithfulness' – a joint probability distribution is said to be *faithful* to a DAG if there are no conditional independence relationships between the variables that do not follow from the directed Markov property.[28] In the special case of models without interactions, more specifically, when for any $u \neq u'$ we have

$$\begin{aligned} &E(Y|U = u, X = x_1) - E(Y|U = u, X = x_2) \\ &\quad = E(Y|U = u', X = x_1) - E(Y|U = u', X = x_2) \end{aligned}$$

a zero causal effect leads to the conclusion that $E(Y|U, X)$ does not depend on X. Such interactions, of course, can never be completely ruled out as U is unobservable. However, one has to presume that even when allowing for possible interactions between X and U, it would be rare in practice to obtain a zero causal effect without at least having that $E(Y|U = u, X = x) = E(Y|U = u)$ because the cancellation discussed above requires a very specific numerical configuration.

It would be convenient if (under conditions 1)–3)) $Y \perp\!\!\!\perp X|U$ if and only if $Y \perp\!\!\!\perp G$. This would imply that if we believed that the conditional distribution of Y under intervention in X is the same as when X is just observed (i.e., the DAG has a causal interpretation so $p(y|x, u) = p(y|do(X = x), u)$) and if we disregard the particular numerical cancellations discussed above, we could test for a causal effect by checking for association between G and Y. Using Equation (2) and integrating out x and u gives the marginal joint distribution of (Y, G)

$$p(y, g) = p(g) \sum_u p(u) \sum_x p(y|u, x)p(x|u, g)$$

We can see that if $Y \perp\!\!\!\perp X|U$, i.e., if $p(y|u, x) = p(y|u)$, then

$$p(y, g) = p(g) \sum_u p(y|u)p(u) \sum_x p(x|u, g) = p(g)p(y)$$

So $Y \perp\!\!\!\perp X|U \Rightarrow Y \perp\!\!\!\perp G$. We note that the joint distribution also factorizes if $p(x|u, g) = p(x|u)$, i.e., if X and the IV are not associated but this has been ruled out by core condition 2).

The reverse argument $Y \perp\!\!\!\perp G \Rightarrow Y \perp\!\!\!\perp X|U$ does not hold, however, even if we know that $p(x|u,g) \neq p(x|u)$. Again, this is due to the possibility of very specific numerical cancellations that might induce a factorization of $p(y,g)$ without the desired independence.

Summarizing, we have

$$G \perp\!\!\!\perp Y \Leftarrow Y \perp\!\!\!\perp X \mid U \Rightarrow p(y \mid do(x)) = p(y) \tag{6}$$

whereas the reverse implications can be violated if specific numerical patterns of the parameters of the involved distributions occur. As the latter would appear unlikely in practice, we will consider it reasonably safe to regard the three statements in Equation (6) as 'equivalent for practical purposes'. The complementary implications of Equation (6), especially $G \not\perp\!\!\!\perp Y \Rightarrow Y \not\perp\!\!\!\perp X|U$ are true in any case, so that if a test finds that G and Y are not independent we can conclude that the outcome does depend on the phenotype given the confounders.

5.2 Retrospective view

In the case control study situation where we condition on the outcome Y, the data only allow us to identify properties of the distribution $p(x,u,g|y)$ which from Equation (2) is seen to be

$$p(x,u,g|y) = \frac{p(y|u,x)p(x|u,g)p(u)p(g)}{\sum_{x,u} p(y|u,x)p(x|u)p(u)}.$$

This will typically not factorize in any way: there are no independencies among X, G, U conditional on Y, as reflected in the moral link induced in the graph by this conditioning and as discussed in Section 4.2. However, assuming, as in the prospective case, that for practical purposes 'no causal effect' is equivalent to $Y \perp\!\!\!\perp X|U$, then the above conditional distribution becomes

$$p(x,u,g|y) = \frac{p(y|u)p(x|u,g)p(u)p(g)}{\sum_{u} p(y|u)p(u)}$$

if there is no causal effect. By summing out x and u we then find that $p(g|y) = p(g)$. Hence, for a case control study, we can also expect that if there is no causal effect there should be no association between Y and G, or equivalently, if we find an association between Y and G then there is a causal effect.

6 Identification of a causal effect

Identifiability of the ACE from observational data requires more than the core conditions 1)–3). We consider the additional assumption that all conditional expectations of the variables in Figure 1 are linear in their graph parents without interactions (Section 6.1). When linearity is doubtful, for instance because the response is binary, it is possible to derive upper and lower bounds for the causal effect (Section 6.2). Section 6.3 addresses

the question of why the non-linear case cannot be treated in a similar manner to the linear additive case.

6.1 The easy case: linearity without interactions

In the following we assume general linear models for the dependencies among the variables Y, X, G and U. Furthermore, we suppose that all dependencies only affect the mean. In other words, we assume that

$$\begin{aligned} E(Y|X = x, U = u) &= \alpha + \beta_1 x + \beta_2 u \\ E(X|G = g, U = u) &= \gamma + \delta_1 g + \delta_2 u \end{aligned} \tag{7}$$

with both X and Y having constant (possibly different) conditional variances. In addition, we assume that the first expectation is the same if we intervene in X, that is, the link is causal and hence

$$E(Y|do(X = x_0), U = u) = \alpha + \beta_1 x_0 + \beta_2 u$$

In this framework, as noted in Section 3.2, β_1 is the causal parameter that we are interested in since $\mathrm{ACE}(x_1, x_2) = \beta_1(x_1 - x_2)$. It cannot be estimated from a linear least squares regression of Y on X and U, as U is unobserved, nor is it estimable from a linear least squares regression of Y on X alone, as X and U are correlated. A linear least squares regression of X on G, however, will yield a consistent estimate $r_{X|G}$ of δ_1, the coefficient of G, because G and U are uncorrelated. Since the regression coefficient $r_{Y|G}$ of G in a linear least squares regression of Y on G alone can be shown to be consistent for $\beta_1\delta_1$, the required causal parameter, β_1, can be consistently estimated as the ratio

$$\hat{\beta}_1 = \frac{r_{Y|G}}{r_{X|G}}$$

(see Appendix for technical details.) Obviously we need $r_{X|G}$ to be non-zero but this is provided by condition 2): $G \not\perp\!\!\!\perp X$. The variance of $\hat{\beta}_1$ depends on the conditional variance of Y given $do(X)$ as well as on the variance of the IV, G, and inversely on the covariance of X and G (cf. discussion by Martens *et al.*[11]). A strong G–X association thus increases the precision of the IV estimator, as mentioned in Section 4.3. Note that these results rely on models without interactions. In particular, there can be no interactions with the unobserved confounder U and, since U is unknown, this is obviously an untestable assumption.

Straightforward generalization to the case where G is binary is possible. In this case, the parameter δ_1 would then be the mean difference in X for the two different values of G. In fact, G can have more than two values as long as its relation with X is linear. The implication for Mendelian randomization applications when G could assume three values, one for each genotype in the simplest diallelic case, is that the expected change in X between genotypes 0 and 1 must be the same as the expected change in X between genotypes 1 and 2. In other words, the genetic model must be additive. There is no sensible genetic model that is consistent with this requirement for a polymorphic locus with more than three genotypes.

The case where X is binary is often tackled in the econometrics literature using a *dummy endogenous variable model*.[7,29] This is based on a threshold approach which assumes an underlying unobservable continuous variable X_c with linear conditional expectation $E(X_c|G,U)$ as given in Equation (7) above. The observable quantity is X where $X = 1$ if $X_c > 0$ and $X = 0$ otherwise. It can be shown that β_1 can still be recovered as before. (See Appendix for details.)

6.2 Bounds on the causal effect

When only the core conditions 1)–3) can be assumed without additional assumptions such as linearity and no interactions, the causal effect is not identifiable and we can at best give lower and upper bounds that will contain the ACE.[13–15] This method, however, requires all observable variables to be binary or categorical which could be achieved by suitable categorization of continuous variables. In the binary case, let $p_{ij\cdot k} = p(Y = i, X = j|G = k)$ represent the conditional probabilities which can be estimated easily from the data using the corresponding relative frequencies. It can be shown that the bounds for the average causal effect, $\mathrm{ACE}(1,0) = p(Y = 1|do(X = 1)) - p(Y = 1|do(X = 0))$, are given by

$$\left.\begin{array}{c} p_{11.1} + p_{00.0} - 1 \\ p_{11.0} + p_{00.1} - 1 \\ p_{11.0} - p_{11.1} - p_{10.1} - p_{01.0} - p_{10.0} \\ p_{11.1} - p_{11.0} - p_{10.0} - p_{01.1} - p_{10.1} \\ -p_{01.1} - p_{10.1} \\ -p_{01.0} - p_{10.0} \\ p_{00.1} - p_{01.1} - p_{10.1} - p_{01.0} - p_{00.0} \\ p_{00.0} - p_{01.0} - p_{10.0} - p_{01.1} - p_{00.1} \end{array}\right\} \leq \mathrm{ACE} \leq \left\{\begin{array}{c} 1 - p_{01.1} - p_{10.0} \\ 1 - p_{01.0} - p_{10.1} \\ -p_{01.0} + p_{01.1} + p_{00.1} + p_{11.0} + p_{00.0} \\ -p_{01.1} + p_{11.1} + p_{00.1} + p_{01.0} + p_{00.0} \\ p_{11.1} + p_{00.1} \\ p_{11.0} + p_{00.0} \\ -p_{10.1} + p_{11.1} + p_{00.1} + p_{11.0} + p_{10.0} \\ -p_{10.0} + p_{11.0} + p_{00.0} + p_{11.1} + p_{10.1} \end{array}\right.$$

These bounds are *sharp* in the sense that they cannot be improved upon without making additional assumptions.[15] Wide bounds, possibly containing zero, will reflect that the data, including the IV, are not very informative for the causal effect. Reasons for this could be small sample size, weak instrument or a true ACE close to zero.

In principle, these bounds can also be calculated for factors with more than two levels. However, the more categories there are, the more difficult the computations and the less informative the bounds. In Mendelian randomization applications, Y is usually binary and G can often be considered as binary with one genotype having an effect on the trait and the others pooled into one that has no effect. To calculate the above bounds, we would dichotomize the usually continuous phenotype X as being above or below a given threshold and compare the bounds for different discretizations of X. Note, however, that the core conditions then have to be satisfied for the dichotomized version of X, which is not automatically implied if they hold for the continuous one.

The above bounds are not applicable to the case–control data situation because they rely on estimating $p_{ij\cdot k}$ which cannot be done when the data have been selected on Y. However, knowledge of the disease prevalence, $p(y)$, will almost always be available.

This, together with the genotype distribution $p(g)$ permits calculation of the bounds for case–control data since

$$p(y, x|g) = p(g, x|y)\frac{p(y)}{p(g)}$$

and $p(g, x|y)$ *can* be estimated from the data.

In practice, the bounds provide an assessment of how informative the data themselves are in the absence of additional parametric assumptions. Any parametric approach should therefore be supplemented by the computation of these bounds. Furthermore, by insisting that all the upper bounds are greater than or equal to the lower bounds, testable constraints arise which, if violated, imply that at least one of the core conditions is not satisfied.

6.3 The difficult case

From Section 6.1 and the corresponding derivations in the Appendix we can see that in order to use the IV technique we need to, firstly, specify what our causal parameter is and, secondly, determine how it relates to the regression parameters of a regression of X on G and a regression of Y on G. The latter involves marginalizing over U and the result in the non-linear case is typically not independent of the unknown distribution of U.

6.3.1 Causal parameter for non-linear models

Assume we have a binary response variable Y, as is the case in many epidemiological applications. The ACE is then the *risk difference*. To determine the ACE, we have to integrate out U in order to determine $E(Y|do(X = x))$, which in the case of a non-linear dependence of Y on X and U will typically depend on the unknown distribution of U. For instance, assuming a logistic regression we obtain

$$E(Y|do(X = x)) = \int \frac{\exp\{\alpha + \beta_1 x + \beta_2 u\}}{1 + \exp\{\alpha + \beta_1 x + \beta_2 u\}} p(u)\, \mathrm{d}u \tag{8}$$

where $p(u)$ is the unknown density of U. The above expectation cannot generally be written in the same model form with a different constant. In particular, for this case,

$$E(Y|do(X = x)) \neq \frac{\exp\{\alpha^* + \beta_1 x\}}{1 + \exp\{\alpha^* + \beta_1 x\}} \tag{9}$$

even when U is assumed to be normally distributed. Greenland *et al.*[30] discuss this problem referring to it as non-collapsibility of the logistic regression model.

Alternative causal parameters are the odds ratio or relative risk. The former is particularly important when the data come from case–control studies, as these only allow identification of odds ratios. We can define the causal odds ratio (COR) as

$$\mathrm{COR}(x_1, x_2) = \frac{p(Y = 1|do(X = x_1))}{p(Y = 0|do(X = x_1))} \frac{p(Y = 0|do(X = x_2))}{p(Y = 1|do(X = x_2))}$$

where $p(Y = 1|do(X = x_1)) = E(Y|do(X = x_1))$. However, from Equations (8) and (9), it is not guaranteed that β_1 is the causal parameter of interest in the sense that

$$\mathrm{COR}(x_1, x_2) \neq \exp\{\beta_1(x_1 - x_2)\}$$

In fact, COR again depends on the unknown confounder distribution. For the causal relative risk (CRR) defined as

$$\mathrm{CRR}(x_1, x_2) = \frac{p(Y = 1|do(X = x_1))}{p(Y = 1|do(X = x_2))}$$

we consider a (causal) log-linear model corresponding to this parameter

$$E(Y|X = x, U = u) = E(Y|do(X = x), U = u) = \exp\{\alpha + \beta_1 x + \beta_2 u\}$$

which does imply that $E(Y|do(X = x)) = \exp\{\alpha^* + \beta_1 x\}$ after marginalizing over U (where typically $\alpha^* \neq \alpha$), independently of the distribution of U. Thus, the log-linear shape is retained, CRR is $\exp\{\beta_1(x_1 - x_2)\}$ and β_1 is the relevant causal parameter.

6.3.2 Relationship between regression parameters

The next step is to investigate how the regression parameters estimated by a regression of Y on G and a regression of X on G are related to the causal parameter of interest. From the Appendix we see that this relies on working out

$$E(Y|G = g) = E_U E_{X|U,G=g} E(Y|X, U)$$

that is, we compute the conditional expectation of Y given (X, U), then integrate out X with regard to its conditional distribution given $(U, G = g)$ and then integrate out U with respect to its marginal distribution, thus exploiting the core conditional independencies to obtain the conditional expectation of Y given G. The derivation in the Appendix crucially depends on the different expectations being additive in the conditioning variables so that taking expectations is straightforward. But if we assume a logistic regression for a binary response variable Y, $E(Y|X, U) = p(Y = 1|X, U)$ is not additive in X and U and so the expectation $E_{X|U,G=g}E(Y|X, U)$ is not straightforward to compute as it involves an integral to which there is not necessarily an analytic solution. Thompson *et al*.[31] suggest an approximation. However, this approximation ignores U and assumes that $Y \perp\!\!\!\perp G|X$ which, if true, would mean that an IV approach is unnecessary as there would be no confounding and the effect of X on Y could be estimated from the data. Hence, this does not provide a way of identifying the COR in the situation where there is confounding but it can be used as a heuristic check for the presence of confounding.[32]

If we model $p(Y = 1|X, U)$ log-linearly corresponding to the CRR, then $E(Y|X, U) = p(Y = 1|X, U)$ is multiplicative in X and U and is again not additive. Hence, just as for the logistic relationship, solving $E(Y|G = g) = E_U E_{X|U,G=g} E(Y|X, U)$ requires an approximation. The derivation provided in Thomas and Conti[33] also ignores U and assumes that $Y \perp\!\!\!\perp G|X$. The same general problem arises with the approximation for the probit link that these authors refer to. The probit model can however be

used under certain assumptions when it is held that all binary variables are generated by unobserved underlying continuous variables that have a joint multivariate normal distribution.[34]

7 Complications for Mendelian randomization

The limitations of Mendelian randomization, from the perspective of complicating features leading to poor estimation of the required genotype–phenotype and genotype–disease associations, have been discussed in detail in several places in the literature.[5,33,35,36] More crucially, biological complications can sometimes violate one or more of the core conditions 1)–3) so that Figure 1 no longer applies. Our focus here is on explicating what any added complexity implies with regard to meeting these conditions. This is illustrated using DAGs that are ideally dictated by the biology.

Linkage disequilibrium refers to the association between alleles at different loci across the population, as is the case when loci are physically close on the chromosome and thus tend to be inherited together, or may be due to other reasons such as natural selection, assortative mating and migration.[37] When our chosen gene G_1 is in linkage disequilibrium with another gene G_2 which has a direct or indirect influence on the disease Y, condition 3) ($Y \perp\!\!\!\perp G_1|(X, U)$) might be violated, as shown in Figure 5(a), or else condition 1) ($G \perp\!\!\!\perp U$) might be violated, as shown in Figure 5(b).

Pleiotropy is the phenomenon whereby a single gene may influence several traits. If the chosen instrument G is associated with another intermediate phenotype which also has an effect on the disease Y (Figure 6(a)), condition 3), $Y \perp\!\!\!\perp G|(X_1, U)$, is again violated if we do not also condition on X_2. Moreover, a genetic polymorphism under study might have pleiotropic effects that influence confounding factors like consumption of tobacco or alcohol, for example, Davey Smith and Ebrahim[5]. This is represented in Figure 6(b) and violates condition 1).

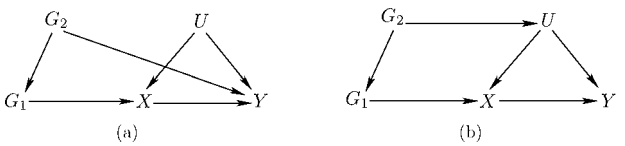

Figure 5 Linkage disequilibrium in a Mendelian randomization application.

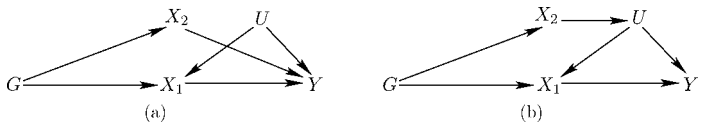

Figure 6 Pleiotropy in a Mendelian randomization application.

Genetic heterogeneity arises when more than one gene affects the phenotype. The core conditions may still hold for the chosen instrument G_1 in the situation of Figure 7, where none of the other genes influence Y in any way other than via their effect on X. If instead the situation is similar to Figure 5(b), for example, the core conditions may be violated as already explained. Also, in Figure 7, genetic heterogeneity could weaken the G_1–X association.

Population stratification denotes the co-existence of different disease rates and allele frequencies within subgroups of individuals. In Figure 8(a), we see that condition 3), $Y \perp\!\!\!\perp G_1 | (X, U)$, is again violated: we need to condition on the population subgroup as well. However, if population stratification causes an association between allele frequencies and phenotype levels, as in Figure 8(b), all conditions for G to be an instrument are still satisfied, and, in this situation, the G–X association may in fact be strengthened or weakened, as a result.

Furthermore, the true biological situation could involve several genes affecting several phenotypes whose joint influence on the disease of interest is subject to confounding (Figure 9). Keavney *et al*.[38] consider six lipid-related genes and two plasma apolipoproteins affecting CHD. Each gene was considered as a separate instrument and the ratio of the plasma lipoproteins taken as the intermediate phenotype. However, the reported genotype–CHD associations were not consistent with the genotype–phenotype associations whereby genes that adversely affected lipoprotein levels were not necessarily associated with increased risk of CHD. One explanation is that the underlying biology is more like Figure 9 and that interactions between genes and between phenotypes need to be taken

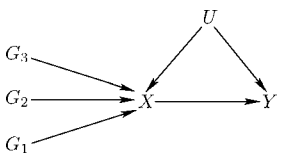

Figure 7 Genetic heterogeneity in a Mendelian randomization application.

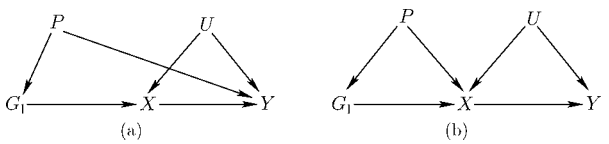

Figure 8 Two examples of population stratification for Mendelian randomization.

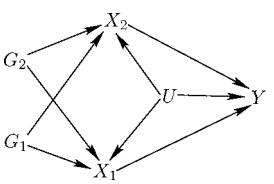

Figure 9 A more complicated example for gene–phenotype–disease relations.

into account. There is currently no simple extension of the current IV approach that is applicable to this situation.

8 Discussion

It seems to us that applications of Mendelian randomization can benefit from a formal framework for causal inference.[39] For example, the 'do' operation of Pearl[9] that we have used in the present paper allows explicit specification of what the causal aim under investigation is and under which conditions it can be attained. We argue and confirm that Mendelian randomization can often be reasonably assumed to satisfy the conditions of an IV. Moreover, these core conditions can be represented using DAGs and can hence be verified by visual inspection (e.g., Section 7). Our focus regarding causal inference is on testing for and estimating the causal effect, and on the special case of retrospective as opposed to prospective studies. Our findings are as follows. Testing for a phenotype–disease causal effect by testing for a genotype–disease association, as suggested by Katan[16], is reasonable for practical purposes (Section 5). For calculation of the ACE, one must rely on additional strong parametric assumptions such as linearity and no interactions but these are not typically justifiable for epidemiological applications with a binary disease outcome. At least, one should carry out sensitivity analyses if such models are used. In the non-linear/interaction case, even the specification of the causal parameter is not obvious and determination of its relationship to the relevant regression parameters is not straightforward (Section 6). 'There is, in fact, no agreed upon generalization of IVs to non-linear systems'[9]. This is particularly alarming as in a case–control study, generally only the odds ratio can sensibly be considered. Assuming no more than the core conditions for a genotype to be an instrument, bounds on the ACE can always be calculated if all variables are binary and these can be adapted to the case–control situation if the disease prevalence is known.

The above framework leads us to see the limitations of Mendelian randomization as falling into two main categories: in some situations, the core assumptions for a genotype to be an instrument are not plausible (Section 7), and in other cases, the parametric assumptions that are required to calculate the causal effect are not reasonable. While there is much discussion in the literature about the difficulty of obtaining good estimates from genetic association studies, in particular, due to lack of power or because the estimates come from different studies,[17] the current approaches to testing and estimating the relevant causal effects are rarely challenged. For instance, the ratio point estimate for the relevant regression parameter which is valid in the additive linear case (Section 6.1) has been advocated as appropriate for binary outcomes,[33,40] which seems dubious to us. Even in a recent article,[41] while it is noted that linearity is required for this estimate, the assertion that such assumptions are 'not easily satisfied in a Mendelian randomization setting' is not discussed.

In summary, we regard the following issues as most important and pressing for further research in order to widen the applicability of the method of Mendelian randomization:

general methods for the non-linear case and consideration of different causal parameters. The econometric literature could be a promising source of IV methods for some non-linear models.[42,43] In particular, proposals for estimation of the CRR (cf. Mullahy[44] or Windmeijer and Santos Silva[45]) have to be scrutinized with regard to the required assumptions and might be combined with logit models when events are rare.[46] A drawback of these methods is that interactions between the phenotype and the confounding factors are excluded which is a restrictive assumption and calls for generalizations as well as thorough sensitivity analyses. Also, advances made in connection with the so-called local causal effect based on randomized trials with partial compliance need to be scrutinized as to whether this causal parameter is meaningful in a Mendelian randomization setting and whether it may be estimated in situations where the ACE cannot. A general model class proposed in this context is the one of structural mean models (cf. overview in Hernán and Robins[12]). Bayesian approaches have also been proposed for this RCT scenario.[47] These, again, require further investigation.

Appendix

Background on the linear and no interactions case

Let us first show why β_1 is the causal parameter of interest. With the assumptions from Section 6.1, we have that

$$\begin{aligned} E(Y|do(X=x)) &= E_{U|do(X=x)}E(Y|do(X=x),U) \\ &= E_U E(Y|do(X=x),U) \\ &= \alpha + \beta_1 x + \beta_2 \mu_U \\ &= \alpha^* + \beta_1 x \end{aligned}$$

where $\mu_U = E(U)$ and using obvious notation for iterated conditional expectation. The second equality holds because $do(X=x)$ is not informative for U as there cannot be any dependence between an intervention and an unknown variable (graphically, intervening on X removes the arrows leading into X from Figure 1). From the above, we obtain

$$\mathrm{ACE}(x_1, x_2) = \beta_1(x_1 - x_2)$$

so we are interested in estimating β_1.

Using the usual conditioning without intervention, a regression of Y on X alone corresponds to

$$\begin{aligned} E(Y|X=x) &= E_{U|X=x}E(Y|X=x,U) \\ &= \alpha + \beta_1 x + \beta_2 \mu_{U|X=x} \end{aligned}$$

where $\mu_{U|X=x} = E(U|X=x)$ is typically not constant in x, in particular not equal to μ_U, due to the dependence between X and U in the non-interventional case, that is, in

the observational regime. Hence β_1 cannot be identified from a regression of Y on X alone.

Instead consider a regression of Y on G alone. This corresponds to

$$\begin{aligned} E(Y|G=g) &= E_{(X,U)|G=g}E(Y|X,U,G=g) \\ &= E_{U|G=g}E_{X|U,G=g}E(Y|X,U) \quad \text{since} \quad Y \perp\!\!\!\perp G|(X,U) \\ &= E_U E_{X|U,G=g}E(Y|X,U) \quad \text{since} \quad U \perp\!\!\!\perp G \\ &= E_U(\alpha + \beta_1(\gamma + \delta_1 g + \delta_2 U) + \beta_2 U) \\ &= \alpha + \beta_1\gamma + \beta_1\delta_1 g + (\beta_1\delta_2 + \beta_2)\mu_U \\ &= \alpha^* + \beta_1\delta_1 g \end{aligned}$$

Hence, the coefficient of G in a regression of Y on G is $\beta_1\delta_1$.

Furthermore, a regression of X on G alone corresponds to

$$\begin{aligned} E(X|G=g) &= E_{U|G=g}E(X|G=g,U) \\ &= E_U E(X|G=g,U) \\ &= \gamma + \delta_1 g + \delta_2\mu_U \end{aligned}$$

so the coefficient of G in this regression is δ_1. Thus, as given in Section 6.1, the causal parameter of interest, β_1, can be estimated from the ratio of these two regression coefficients. This is usually done in a 'two-stage-least-squares' procedure: more details can be found in Stewart and Gill[48].

Threshold models

The above can be generalized to the case of binary G and X in the following way. Let $E(Y|X=x,U=u) = \alpha + \beta_1 x + \beta_2 u$, as before. Consider an underlying unobserved variable X_c with

$$E(X_c|G=g,U=u) = \gamma + \delta_1 g + \delta_2 u$$

and define

$$X = \begin{cases} 1, & X_c > 0 \\ 0, & \text{otherwise} \end{cases}$$

The dependence structure can be represented as in Figure 10 where the relationship between X_c and X is deterministic. The conditional independencies with respect to *all* variables are $Y \perp\!\!\!\perp (G, X_c)|(U,X)$, $X \perp\!\!\!\perp (G,U)|X_c$ and $G \perp\!\!\!\perp U$. But as X_c is not observed we have $Y \perp\!\!\!\perp G|(U,X)$, $X \not\perp\!\!\!\perp G$ and still $G \perp\!\!\!\perp U$ for the remaining variables. Hence, the core conditions 1)–3) apply to (G,U,X,Y), ignoring X_c, and β_1 is still the parameter we are interested in as it describes the required causal effect.

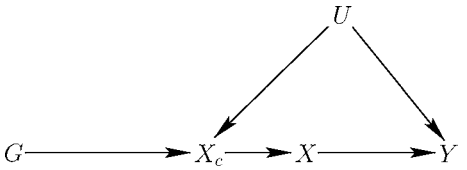

Figure 10 Mendelian randomization for a threshold model.

By an argument similar to that presented above, we have (assuming $\delta_2 > 0$)

$$\begin{aligned} E(Y|G=g) &= E_U E_{X|U,G=g} E(Y|X,U) \\ &= E_U(\alpha + \beta_1 I(\gamma + \delta_1 g + \delta_2 U > 0) + \beta_2 U) \\ &= \alpha + \beta_1 P_U(\gamma + \delta_1 g + \delta_2 U > 0) + \beta_2 \mu_U \\ &= \alpha + \beta_1 P_U \frac{(U > (-\gamma - \delta_1 g))}{\delta_2} + \beta_2 \mu_U \end{aligned}$$

Recall that G is binary and assumes the value 0 or 1. If we let $\xi_0 = -\gamma/\delta_2$ and $\xi_1 = (-\gamma - \delta_1)/\delta_2$, the above model can be written as

$$E(Y|G=g) = \alpha + \beta_1 P_U(U > \xi_0) + \beta_1(P_U(U > \xi_1) - P_U(U > \xi_0))g + \beta_2 \mu_U$$

This is linear in G and its coefficient $r_{Y|G=1} = \beta_1(P_U(U > \xi_1) - P_U(U > \xi_0))$ can be estimated by linear least squares. The relationship between the observable X and G works out to be

$$\begin{aligned} E(X|G=g) &= E_U E(X|G=g,U) \\ &= P_U(\gamma + \delta_1 g + \delta_2 U > 0) \\ &= P_U(U > \xi_0) + (P_U(U > \xi_1) - P_U(U > \xi_0))g \end{aligned}$$

which is also linear in G. The relevant coefficient $(P_U(U > \xi_1) - P_U(U > \xi_0))$ can again be estimated by linear least squares. Note that the modelling assumptions made here, in particular, about the role of U, are very specific and cannot fully be tested empirically.

Acknowledgements

The authors acknowledge research support from Leverhulme Research Interchange Grant F/07134/K, which instigated our collaboration, and from the Centre for Advanced Study at the Norwegian Academy of Science and Letters. We are also grateful to Steffen Lauritzen and Phil Dawid for helpful comments and discussions.

References

1 Fisher RA. *The design of experiments*, first edition. Oliver & Boyd, 1926.
2 Fisher RA. *The design of experiments*, eigth edition. Hafner, 1970.
3 Peto R, Doll R, Buckley JD, Sporn MB. Can dietary beta-carotene materially reduce human cancer rates? *Nature* 1981; **290**: 201–8.
4 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* 1994; **330**: 1029–35.
5 Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**: 1–22.
6 Katan MB. Commentary: Mendelian randomization, 18 years on. *International Journal of Epidemiology* 2004; **33**: 10–11.
7 Bowden RJ, Turkington DA. *Instrumental variables*. Cambridge University Press, 1984.
8 Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**(434):444–55.
9 Pearl J. *Causality* (Chapter 7). Cambridge University Press, 2000.
10 Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**:722–9.
11 Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables, application and limitations. *Epidemiology* 2006; **17**: 260–7.
12 Hernán MA, Robins JM. Instruments for causal inference. An epidemiologist's dream? *Epidemiology* 2006; **17**: 360–2.
13 Robins JM. The analysis of randomized and nonrandomized aids treatment trials using a new approach to causal inference in longitudinal studies. In L, Sechrest H, Freeman and A. Mulley eds. *Health Service Research Methodology. A Focus on AIDS*. U.S. Public Health Service, 1989: 113–59.
14 Manski CF. Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings* 1990; **80**: 319–23.
15 Balke AA, Pearl J. Counterfactual probabilities: computational methods, bounds and applications. In Mantaras RL, Poole D. eds. *Proceedings of the 10th Conference on Uncertainty in Artificial Inteligence*. 1994: 46–54.
16 Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986; **i**: 507–8.
17 Minelli C, Thompson JR, Tobin MD, Abrams KR. An integrated approach to the Meta-Analysis of genetic association studies using Mendelian randomization. *American Journal of Epidemiology* 2004; **160**: 445–52.
18 Scandinavian Simvastin Survival Study (4S). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease. *Lancet* 1994; **344**: 1383–9.
19 Heart Protection Study Collaborative Group. MRC/BHF heart protection study of cholesterol lowering with simvastin in 20,536 high-risk individuals. *Lancet* 2002; **360**: 7–22.
20 Pearl J. Causal diagrams for empirical research. *Biometrika* 1995; **82**: 669–710.
21 Lauritzen SL. Causal inference from graphical models. In Barndorff-Nielsen OE, Cox DR, Kluppelberg C. eds. *Complex Stochastic Systems*. Chapter 2, Chapman & Hall, 2000: 63–107.
22 Dawid AP. Influence diagrams for causal modelling and inference. *International Statistical Review* 2002; **70**: 161–89.
23 Dawid AP. Causal inference using influence diagrams: the problem of partial compliance. In Green PJ, Hjort NL, Richardson S. eds. *Highly Structured Stochastic Systems*. Oxford University Press, 2003: 45–81.
24 Goldberger AS. Structural equation methods in the social sciences. *Econometrica* 1972; **40**: 979–1001.
25 Pearl J. *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann, 1988.
26 Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. *Probabilistic networks and expert systems*. Statistics for Engineering and Information Science, Springer-Verlag, 1999.
27 Dawid AP. Conditional independence in statistical theory (with Discussion). *Journal of the Royal Statistical Society, Series B* 1979; **41**: 1–31.
28 Spirtes P, Glymour C, Scheines R. *Causation, prediction and search*, second edition. MIT Press, 2000.

29 Maddala GS. *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press, 1983.

30 Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science* 1999; **14**: 29–46.

31 Thompson JR, Tobin MD, Minelli C. On the accuracy of the effect of phenotype on disease derived from Mendelian randomisation studies. Genetic Epidemiology Technical Report 2003/GE1, Centre for Biostatistics, Department of Health Sciences, University of Leicester, 2003. Retrieved from http://www.prw.le.ac.uk/research/HCG/getechrep.html.

32 Didelez V, Sheehan NA. Mendelian randomisation and instrumental variables: what can and what can't be done. Technical Report 05-02, Department of Health Sciences, University of Leicester, 2005. Retrieved from http://www.homepages.ucl.ac.uk/ucakvdi/vlon.html.

33 Thomas DC, Conti DV. Commentary: the concept of 'Mendelian randomization'. *International Journal of Epidemiology* 2004; **33**: 21–5.

34 Stanghellini E, Wermuth N. On identification of path analysis models with one hidden variable. *Biometrika* 2005; **92**: 337–50.

35 Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* 2004; **33**: 30–42.

36 Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* 2005a; **366**: 1484–98.

37 Lynch M, Walsh B. *Genetics and analysis of quantitative traits*. Sinauer Associates Inc., 1998.

38 Keavney BD, Palmer A, Parish S, Clark S, Youngman LD, Danesh J, McKenzie C, Delphine M, Lathrop M, Peto R, Collins R. Lipid-related genes and myocardial infarcion in 4685 cases and 3460 controls: discrepancies between genotype, blood lipid concentrations, and coronary disease risk. *International Journal of Epidemiology* 2004; **33**: 1002–13.

39 Didelez V, Sheehan NA. Mendelian randomisation: why epidemiology needs a formal language for causality. In Russo F, and Williamson J, ed. *Causality and Probability in the Sciences*. London College Publications, Texts in philosophy 2007; **5**: 263–292.

40 Davey Smith G, Lawlor DA, Harbord R, Timpson N, Rumley A, Lowe GDO, Day INM, Ebrahim S. Association of C-reactive protein with blood pressure and hypertension: life course confounding and Mendelian randomisation tests of causality. *Arteriosclerosis, Thrombosis and Vascular Biology* 2005b; **25**: 1051–56.

41 Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, Leon DA. Limits to causal inference based on Mendelian randomization: a comparison with randomised controlled trials. *American Journal of Epidemiology* 2006; **163**: 397–403.

42 Amemiya T. The nonlinear two-stage least-squares estimator. *Journal of Econometrics* 1974; **2**: 105–10.

43 Hansen LP, Singleton RJ. Generalized instrumental variable estimation of non-linear rational expectation models. *Econometrica* 1982; **50**: 1269–86.

44 Mullahy J. Instrumental variable estimation of Poisson regression models: application to models of cigarette smoking behaviour. *Review of Economics and Statistics* 1997; **79**: 586–93.

45 Windmeijer F, Santos Silva JMC. Endogeneity in count data models: an application to demand for health care. *Journal of Applied Econometrics* 1997; **12**: 281–94.

46 Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology* 1990; **132**: 734–45.

47 Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics* 1997; **25**: 305–27.

48 Stewart J, Gill L. *Econometrics*, second edition. Prentice & Hall, 1998.