Mendelian Randomisation: Why Epidemiology needs a Formal Language for Causality

Vanessa Didelez and Nuala A. Sheehan

ABSTRACT. For ethical or practical reasons, randomised cotrolled trials are not always an option to test epidemiological hypotheses. Epidemiologists are consequently faced with the problem of how to make causal inferences from observational data, particularly when confounding is present and not fully understood. The method of instrumental variables can be exploited for this purpose in a process known as Mendelian randomisation. However, the approach has not been developed to deal satisfactorily with a binary outcome variable in the presence of confounding. This has not been properly understood in the medical literature. We show that by defining the problem using a formal causal language, the difficulties can be identified and misinterpretations avoided.

1 Introduction

Detection and assessment of the effect of some modifiable risk factor on a disease with view to informing public health intervention policies are of fundamental concern in epidemiology. For example, it is now well established that the risk of neural tube defects, such as spina bifida, can be greatly reduced by periconceptual maternal folate supplementation (MRC Vitamin Study Research Group, 1991; Czeizel and Dudás, 1992; Scholl and Johnson, 2000). A simple public health intervention of adding folic acid to flour and bread has been reported to have reduced the risk by 30-50% in the USA and Canada. The House of Commons Hansard Debates of October 19 2005 recommended this strategy for the UK which, if implemented, would be the first mandatory fortification of food in the UK since the compulsory addition of calcium, iron and vitamins B1 and B2 to flour after the second World War.

Clearly, it is important to have solid evidence that such a public health intervention will have an effect. The problem faced by epidemiologists is that an observed association or correlation between a risk factor and a disease

does not necessarily mean that the risk factor is *causal* for the disease, and if the relationship is not causal, the prescribed intervention will be useless. Inferring causality from observational data is difficult as it is not always clear which of two associated variables is the cause, or which the effect. For example, sick people may change their diets or other aspects of their lifestyle (reverse causation). On the other hand, both disease and exposure levels may be associated purely through another possibly unmeasured factor such as smoking (confounding). In randomised controlled trials (RCTs), the random assignment of "treatment" levels to "experimental units" (Fisher, 1926) essentially renders reverse causation and confounding implausible, but such trials are neither ethical nor practical for many exposures of epidemiological interest like smoking, exercise regimes and alcohol consumption, to name but a few (though some attempts at such trials have been made but inevitably suffer from compliance and other problems).

In situations where randomisation *is* possible, epidemiological studies have been severely criticised for the large numbers of reported associations that have been interpreted as causal and have failed to be replicated in large-scale follow-up RCTs. For example, early observational findings suggesting that increased dietary intake of the anti-oxidant vitamin beta-carotene reduces the risk of smoking-related cancers (Peto et al., 1981) were negated by subsequent RCT findings. (Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group, 1994). Since only candidate causes with the strongest observational support are evaluated in RCTs, we can only assume that many reported associations, as yet untested, are even less likely to be causal (Davey Smith et al., 2005). Confounding is usually the main reason for such spurious findings as reverse causation can often be ruled out by the underlying biology. It is sometimes possible to control for confounding but in general it is difficult to know whether all the relevant confounders (or a sufficient subset of these) have been accounted for. Furthermore, confounding in these applications is usually due to social, behavioural or physiological factors which are difficult to control for and particularly difficult to measure accurately. Epidemiological exposures are also prone to reporting bias. Heavy drinkers, for example, will often under-estimate their alcohol intake.

There is hence a need to infer causality from observational data in the presence of confounding that cannot be controlled for because it is not fully understood. A possible approach in this situation is based on the method of instrumental variables (Bowden and Turkington, 1984; Angrist et al., 1996; Greenland, 2000; Pearl, 2000) which is known under the name of *'Mendelian randomisation'* [1] if the instrument is a genetic predisposition (Davey Smith

[1] The term 'Mendelian randomisation' seems to have become a fixed expression in the

and Ebrahim, 2003; Katan, 2004; Thomas and Conti, 2004). For example, observational studies have indicated that elevated plasma homocysteine levels are associated with increased risk of coronary heart disease (CHD) (Ford et al., 2002) but this effect is suspected to be heavily confounded by the usual factors such as smoking and socioeconomic status (Davey Smith and Ebrahim, 2003). RCTs have confirmed that homocysteine levels can be reduced substantially with a small increase in folate consumption (Homocysteine Lowering Trialists' Collaboration, 1998). However, the T allele of the MTHFR gene is known to be associated with higher homocysteine levels than the more common C allele and thus mimics the effect of low folate intake. In the absence of a definitive folate trial, the causal effect of homocysteine levels (and hence of folate intake) on CHD can be investigated by examining the association of the MTHFR genotype with CHD instead. The former association is affected by confounding but the latter can often be assumed to be free of confounding since alleles are assigned randomly from the two copies of the parents and so causality can be inferred. If the relationship between homocysteine and CHD is truly causal, adding a given quantity of folate to flour would also be worthwhile as a public health intervention to reduce CHD risk in the general population.

The practical difficulties typically encountered when inferring causation from observational data are compounded by the theoretical problem of expressing causal aims and methods in a mathematical language. Causal vocabulary features regularly in the epidemiological literature but this is often accompanied by standard regression methods that do not justify any causal conclusions. Despite recently proposed advances towards a formal causal framework for epidemiological applications (Greenland et al., 1999; Robins, 2001; Hernán, 2004; Hernán et al., 2004) such frameworks are not very widely adopted in general and in particular, are not reflected in the Mendelian randomisation literature at all.

The purpose of this article is to show that a formal, mathematically precise, causal framework is required for Mendelian randomisation applications. It is necessary, firstly, to state precisely what the quantity (parameter) of interest is (e.g. the amount by which CHD risk is reduced from adding folate to flour) and secondly, to formalise how associational findings and causal implications are related in order to obtain an estimate for this particular parameter. Failure to adopt a formal approach has led to misconceptions in the medical literature.

---

literature, but note that this is not a randomisation by study design and hence not fully comparable to a RCT. It has been suggested that 'Mendelian deconfounding' would be a better term (Tobin et al., 2004).

2 A Formal Language for Causality

The medical literature often employs causal vocabulary loosely to express something that is more than association between potential risk factors and their effects. Underlying knowledge about the biology of the problem may indeed allow one to deduce the direction of an observed association and "causal pathways" for disease are familiar terms in the epidemiological literature (see Stanley et al. (2000) for example). The central argument in the present paper is that it is imperative to formally differentiate, with appropriate mathematical notation, between association and causation in order to be explicit about what can be inferred about causality from an observational study. Even the term "causal effect" is used loosely in practice and can mean different things in different settings. We present three approaches to defining a formal language for causality, each of which uses specific mathematical notation to represent that we are interested in interventions such as the public health intervention of adding folic acid (folate) to flour.

2.1 Interventions

As in Pearl (1995); Lauritzen (2000); Dawid (2002, 2003) we regard causal inference to be about the effect of intervening in a given system. For the applications we are considering, this would typically be the motivation for investigating a causal effect. There are many other notions of causality such as used in a courtroom for retrospective assignment of guilt, but we will not consider any other interpretations here. Let X be the cause under investigation and Y the response. In epidemiological applications, X would be the intermediate phenotype (homocysteine level) and and Y would be the disease status (CHD). We focus on the question of whether intervening on X has an effect on Y. 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 remaining variables in the system other than through the resulting changes in X. This is clearly an idealistic situation and not always easily justified for the examples of public health interventions given above. For example, increasing dietary folate will not determine a specific homocysteine level. However, a causal analysis can be used to generate hypotheses that can afterwards be investigated by controlled randomised trials where applicable. Moreover, if a risk factor is found to be causal in the above sense, different types of intervention can then be explored.

2.2 Three Definitions of Causal effect

Roughly speaking, the causal effect contrasts the effects of different interventions in X on the outcome Y, in some sense or another. We will now present three different approaches to doing this.

Pearl's $do(\cdot)$ Operator

Pearl (2000) suggests the notation $P(Y|do(X = x))$ to distinguish between conditioning on intervention in X and 'ordinary' observational conditioning $P(Y|X = x)$ which is sometimes denoted by $P(Y|see(X = x))$ to make the distinction clearer. The former reflects how the distribution of Y should be modified given the information that X has been 'forced' to take on the value x by some external intervention, whereas the latter reflects how the distribution of Y should be modified when we have simply observed $X = x$. The average causal effect (ACE) is then defined as the difference in expectations under different settings of X:

$$ACE(x_1, x_2) = E(Y|do(X = x_1)) - E(Y|do(X = x_2)) \tag{1.1}$$

where x_2 is often chosen to represent some baseline value. In particular, X is regarded as causal for Y if the average causal effect (1.1) is non-zero for some values x_1, x_2 with $x_1 \neq x_2$.

Regime Indicators

This approach goes back to Pearl (1993) and has been further advocated by Lauritzen (2000); Dawid (2002, 2003); Dawid and Didelez (2005). It is based on an indicator F_X assuming values in $\mathcal{X} \cup \emptyset$ with $F_X = x$ if X is being set to the value x by external intervention and $F_X = \emptyset$ (or 'F_X is idle') indicates that X is allowed to arise 'naturally'. Observe that

$$P(Y|do(X = x)) = P(Y|X = x, F_X = x)$$

and

$$P(Y|see(X = x)) = P(Y|X = x, F_X = \emptyset).$$

Due to the deterministic relationship between $F_X = x$ and X we have that $P(Y|X = x, F_X = x) = P(Y|F_X = x)$, or that X is independent of any other variable given $F_X = x$.[2] The advantage of including an intervention indicator like F_X is that the intervention is made explicit and, as we will see later, can be represented visually in a graph. Besides, while the properties of the $do(\cdot)$ operator need to be formulated in separate 'axioms' (see Pearl, 2000, Section 3.4), F_X can be treated as a decision variable that is conditioned on and the properties of conditional independence can be applied (Dawid, 1979) to yield Pearl's axioms. In this context it should be noted that because F_X is a decision variable, it must always be in the conditioning set of a conditional probability, and when we write $Y \perp\!\!\!\perp F_X|X$, for example,

[2] Note that regime indicators can also be used for non–deterministic, i.e. random regimes, where the value to be assigned is drawn from a distribution. We do not go into more details here, but see Dawid and Didelez (2005).

we mean that $P(Y|X = x, F_X = \emptyset) = P(Y|X = x, F_X = x)$ for all $x \in \mathcal{X}$. This intervention variable also permits generalisation of causal inference to other types of intervention. For instance, interventions in X that depend on variables that have been observed before the intervention took place can be considered to reflect, for example, that the dosage of a drug should be different for different sexes and age groups. In such more general situations the type of intervention would be specified via the conditional distribution $P(X|C, F_X)$ where C represents some covariates that are taken into account by the intervention. Conditional interventions such as these are difficult to describe with the $do(\cdot)$ operator.

It is straightforward to reformulate the ACE using the regime indicator:

$$ACE(x_1, x_2) = E(Y|F_X = x_1) - E(Y|F_X = x_2). \tag{1.2}$$

Counterfactuals

A philosophically quite different approach to causality is based on counterfactual variables (Neyman, 1923; Rubin, 1974, 1978; Robins, 1989) Here, Y_{x_1} represents the outcome if a subject is set to the value $X = x_1$ whereas Y_{x_2} is the outcome if the same subject is set to the value $X = x_2$. The variables are counterfactual because they can never both be observed together. With this notation one may define the *individual causal effect* (ICE) as

$$ICE(x_1, x_2) = Y_{x_1} - Y_{x_2}. \tag{1.3}$$

The ACE is now expressed as

$$ACE(x_1, x_2) \quad = \quad E(Y_{x_1} - Y_{x_2}). \tag{1.4}$$

Since this can be rewritten as $E(Y_{x_1}) - E(Y_{x_2})$ we can see that under certain assumptions we do not need to observe Y_{x_1} and Y_{x_2} together in order to make inference about the average causal effect.

In epidemiological applications, the ICE would represent the difference, say, in CHD status of an individual who starts off with high homocysteine levels as opposed to the same individual starting off with low levels, were both observable at the same time. This example demonstrates that, even though it might appear that counterfactuals do not require explicit specification of an intervention, they are only well defined when an intervention is implicit. How could the homocysteine level of a given individual be different from what it actually was? (See Hernán (2004) for further discussion.) Ideally, the risk factor (or potential cause), X, should be a variable on which subjects could be randomised (Rubin, 1974). Hence, all three approaches are only meaningful in the context of interventions that are actually feasible. The implications, for our purposes, are that some concrete public health intervention should be aimed at.

Comments

One could argue informally that the intervention distribution $P(Y|do(X = x))$, or $P(Y|F_X = x)$, corresponds to the distribution of the counterfactual Y_x, but we recommend some caution against this as Y under $do(X = x)$ is still not the same thing as Y_x. In the case of the former, $P(Y|F_X = x)$, we consider *one* variable Y describing the outcome of interest, the distribution of which has to be modified according to what we condition on. In the case of the counterfactual, Y_x, we consider *a number* of potential outcome variables in parallel that have possibly different distributions and, in particular, have a joint distribution. In addition, we note that the ICE being a comparison between two *values* rather than two distributions, does not have a counterpart in the other two frameworks. We will not give any further consideration to the ICE here as it represents a quantity that can never be observed, even in principle. Methods that claim to identify the ICE typically make strong untestable assumptions. Furthermore, it would seem that while the ACE is of obvious public health interest, the ICE might be more of medical/clinical interest. However, our aim is not to discuss the advantages and disadvantages of the three approaches. (For further discussion, see Dawid (2000) and the discussion of that paper, and also Dawid (2007) in this volume.) We wish, rather, to point out that such formal frameworks for causal inference exist and should be used more widely in epidemiological reasoning, especially for Mendelian randomisation applications. We would like to emphasise that it is neither possible to express the desired aim nor quantify the effect of a public health intervention with the "usual" conditional probabilities as the intervention creates a situation that is different from the purely observational one. Some additional 'ingredient' is required to clarify that a public health intervention will be applied to everyone (in a targeted population) and that inferences about this new situation are of interest. Only if this distinction is clearly made, can we work out the precise conditions that will allow such inference from the available data. We will mainly use the approach that includes an intervention indicator F_X but will also demonstrate how some things can be expressed in the other frameworks.

2.3 Identifiability

A causal parameter is identifiable if we can show that it can be estimated consistently from data under the conditions of how those data were obtained (e.g. randomised trial, case-control study, cohort study etc.). Mathematically, this amounts to being able to express the parameter in terms that do not involve the intervention (i.e. without $do(\cdot)$, F_X or counterfactuals) by using 'observational' terms only. These can then be estimated from

data. As noted earlier and detailed in the following section, the distribution under intervention is not necessarily the same as the observational distribution because of confounding, for example. In cases when confounding is well understood, it can be shown that the intervention distribution can mathematically be re-expressed in observational terms and can hence be estimated from the observed data by adjusting for certain variables (Section 3.2) (Pearl, 1995, 2000; Lauritzen, 2000; Dawid, 2002). The instrumental variable technique on which Mendelian randomisation is based, permits a different way of identifying causal parameters when the confounding is poorly understood.

3 The Issue of Confounding

Inferring causality from observational data is complicated by problems that may induce different types of bias. Here, we focus on the problem of confounding, as this is what Mendelian randomisation attempts to circumvent, and show how at least some of the concepts can be formally clarified.

3.1 What is Confounding?

We will not attempt to provide a formal definition of confounding here. This issue is addressed in almost every textbook on epidemiology with varying degree of clarity (see Rothman and Greenland (1998), for example, and for a discussion within the framework of causal graphs see Pearl (2000), chapter 6). We will, however, highlight a few central aspects that will be relevant later.

Confounding could be said to be present whenever $P(Y|X = x, F_X = \emptyset)$ is not equal to $P(Y|F_X = x)$ or similarly if $P(Y|X = x)$ is not $P(Y_x)$. This dual notation reflects the common phrase "correlation is not causation". The well known implication is that a typical model for the regression of Y on X does not necessarily give us any information about the ACE. However, this problem could also be due to reverse causation or time trends which are typically not regarded as confounding but as separate mechanisms inducing bias. A common explanation of confounding is that there exists a variable (or set of variables) C that 'affects' both X and Y. As is well understood in epidemiology (Weinberg, 1993), a crucial implication is that C is not "on the causal pathway" between X and Y. This is important since we do not want to adjust for such variables as the true effect of X on Y could be hidden. Apart from being quite vague, such a requirement is difficult if not impossible to verify from observational data and conditional (in)dependencies since no testable implications arise in either scenario and all variables could be mutually dependent.

Our causal framework allows a formal definition of the requirement that

C should not be "on the causal pathway" as follows:

$$C \perp\!\!\!\perp F_X \tag{1.5}$$

i.e. C is not affected by whether or not we intervene in X. This amounts to saying that we expect the marginal distribution of C to be the same in an observational study as in an RCT, for instance, where X has been randomly allocated. In the counterfactual framework the condition analogous to (1.5) is that there is no counterfactual version of C (i.e. there are no C_{x_1} and C_{x_2}), which implies that it is not affected by an intervention in X (see Dawid's contribution in this volume).

It is perhaps helpful at this point to demonstrate why $P(Y|X = x, F_X = \emptyset)$ and $P(Y|F_X = x)$ are not necessarily the same in the presence of confounding — and with the notion of interventions this can easily be formalised. Consider the above situation where C satisfies (1.5). Since

$$P(Y, X, C|F_X) = P(Y|X, C, F_X)P(X|C, F_X)P(C|F_X)$$

we have from (1.5) that $P(C|F_X) = P(C)$. Furthermore, under intervention, $F_X = x$, we have that $P(X = x'|C, F_X = x) = \mathbf{1}\{x' = x\}$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Hence

$$\begin{aligned} P(Y, X = x, C|F_X = x) &= P(Y|X = x, C, F_X = x)P(C), \text{ and} \\ P(Y, X = x', C|F_X = x) &= 0, \text{ whenever } x' \neq x. \end{aligned}$$

On the other hand, under the observational regime, $F_X = \emptyset$,

$$P(Y, X, C|F_X = \emptyset) = P(Y|X, C, F_X = \emptyset)P(X|C, F_X = \emptyset)P(C).$$

Even if we are willing to assume that $P(Y|X = x, C, F_X = x) = P(Y|X = x, C, F_X = \emptyset)$ (i.e. that $Y \perp\!\!\!\perp F_X|(X, C)$), we can see immediately that the two expressions differ by the factor $P(X|C, F_X = \emptyset)$. This factor reflects that in the observational case, X is informative for C whereas in the intervention case it is not, and this can induce bias if C is predictive for Y. This is crucial to the understanding of the difference between the intervention situation that we are interested in for causal inference, and the observational situation that the data represent.

3.2 Adjusting for Confounding

Confounding can be described in a purely operational manner by showing how one can adjust, or control, for it. Consider a variable, or set of variables C, such that

(i) $C \perp\!\!\!\perp F_X$, i.e. C is "not on the causal pathway", and

(ii) $Y \perp\!\!\!\perp F_X | (X, C)$, i.e. once we know C and X the distribution of Y is independent of how X was generated.

Then C is called 'a sufficient set of covariates' for identifying the ACE (Dawid, in this volume, calls it an 'unconfounder'). More precisely, we can show the following:

$$\begin{aligned} P(Y = y | F_X = x) &= \sum_c P(Y = y | C = c, F_X = x) P(C = c | F_X = x) \\ &\overset{(i)}{=} \sum_c P(Y = y | X = x, C = c, F_X = x) P(C = c | F_X = \emptyset) \\ &\overset{(ii)}{=} \sum_c P(Y = y | X = x, C = c, F_X = \emptyset) P(C = c | F_X = \emptyset) \quad (1.6) \end{aligned}$$

(Pearl, 1995; Dawid, 2002). Now all quantities are observational and the ACE can be calculated by substituting x_1 and x_2, i.e. the causal effect is identifiable from data on X, C, Y on imposition of an additional positivity condition (Dawid, 2002).

Within the counterfactual framework, the corresponding assumptions are:

(i*) $X \perp\!\!\!\perp Y_x | C$, the "no–unmeasured confounder" assumption, and

(ii*) $Y_x = Y$ if $X = x$, the "consistency" assumption.

The distribution of Y_x can then be identified as

$$\begin{aligned} P(Y_x = y) &= \sum_c P(Y_x = y | C = c) P(C = c) \\ &\overset{(i^*)}{=} \sum_c P(Y_x = y | X = x, C = c) P(C = c) \\ &\overset{(ii^*)}{=} \sum_c P(Y = y | X = x, C = c) P(C = c), \end{aligned}$$

where again, all distributions in the last line are 'counterfactual–free' and can hence be estimated from observational data.

Comments

We would like to point out that it is important to state explicitly the conditions that enable adjustment for confounding and that, as before, this requires a formal framework for interventions: otherwise it is difficult to

express why the adjustment (1.6) gives us a desirable quantity. The conditions (i) and (ii) (or (i*) and (ii*)) state precisely what is required to connect the observational data situation to the interventional situation that we are actually interested in. If we do adjust for confounding in the above way we have to justify these conditions based on background knowledge or prior studies for any given data situation. In addition, we need to be able to identify, observe and measure a sufficient set C in a way that ensures that these conditions are satisfied. In practice this may be very difficult; there are many ways to measure smoking behaviour or alcohol intake, for example, and such factors are prone to measurement error and recording bias.

3.3 Confounding in Linear Models

We now address a very popular class of models, linear models without interactions. The assumption of such a model class means that very strong restrictions regarding the shape of (causal) relations are imposed. In practice, these are often not justifiable, but some of the basic concepts explained so far can be illustrated within this simplistic setting. In the following, omission of F_X from the conditioning set implies $F_X = \emptyset$.

Assume that Y is continuous and that the causal dependence of Y on X is linear (possibly after suitable transformations), i.e.

$$E(Y|F_X = x) = \alpha + \beta x.$$

In this case, the average causal effect is $\beta(x_1 - x_2)$ and can be summarised simply by β which is now interpreted as the average effect on Y of increasing X by one unit through some intervention. Similarly, we can make the assumption that $E(Y_x) = \alpha + \beta x$ and obtain the same ACE in the counterfactual framework. In contrast, a standard linear regression models

$$E(Y|X = x, F_X = \emptyset) = \tilde{\alpha} + \tilde{\beta} x$$

and there is a priori no reason to assume that $\tilde{\beta} = \beta$ as discussed in previous sections. In rare cases one may be able to justify $Y \perp\!\!\!\perp F_X \mid X$ implying $E(Y|X = x, F_X = \emptyset) = E(Y|X = x, F_X = x)$ so that in this case $\tilde{\beta} = \beta$ holds. However, this relation usually cannot be assumed in the presence of confounding and the following adjustment is necessary.

Now assume the situation where we include an additional variable (or set of variables) C to adjust for confounding, i.e. C is not on the causal pathway in the sense of (i) and also satisfies (ii). We still assume linear models for all (observational) relationships:

$$E(Y|X = x, C = c) \quad = \quad \alpha + \beta_1 x + \beta_2 c \quad \text{and}$$

$$E(X|C=c) \quad = \quad \gamma + \delta c, \tag{1.7}$$

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:

$$E(Y|C=c, F_X=x) = \alpha + \beta_1 x + \beta_2 c.$$

This reflects assumption (ii): given X and C, the distribution of Y, and hence its expectation, is the same regardless of how X was generated. From the above, we have that

$$\begin{aligned} E(Y|F_X=x) &= E_{C|F_X=x} E(Y|C, F_X=x) \\ &\overset{(i)}{=} E_C E(Y|C, F_X=x) \\ &\overset{(ii)}{=} \alpha + \beta_1 x + \beta_2 \mu_C \\ &= \alpha^* + \beta_1 x, \end{aligned}$$

where $\mu_C = E(C)$ and using obvious notation for iterated conditional expectation. Hence

$$ACE(x_1, x_2) = \beta_1(x_1 - x_2)$$

and so β_1 is the causal parameter of interest.

A regression of Y on X alone corresponds to

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

where $\mu_{C|x} = E(C|X=x, F_X=\emptyset)$ is typically not constant in x and, in particular, is not equal to μ_C due to the dependence between X and C in the observational regime. Hence β_1 cannot be identified from a regression of Y on X alone. However, as we have assumed that C is sufficient for adjustment we can use our adjustment formula (1.6) to obtain

$$\begin{aligned} E(Y|F_X=x) &= \sum_c E(Y|X=x, C=c) P(C=c) \\ &= \sum_c (\alpha + \beta_1 x + \beta_2 c) P(C=c) \\ &= \alpha + \beta_1 x + \beta_2 \mu_C \end{aligned}$$

as desired. Hence, if we have data on X, Y and C we can estimate β_1 from a linear regression of Y on X and C.

Note that if X is binary the ACE is unique (up to its sign) given by $ACE(1,0)$ but in the more general cases of more than two categories of X and/or nonlinear dependency the average causal effect is not necessarily summarised by a single parameter.

4 Formal Graphical Representation

It is useful to introduce a formal graphical representation so that existing substantive background knowledge can be formally encoded and conditions such as (i) and (ii) can be verified visually. We use directed acyclic graphs (DAGs) to represent conditional independencies among a set of joint variables in the following way. A graph is denoted by $\mathcal{G}$ and consists of nodes and directed edges. Every node of the graph represents a variable and these can be linked by directed edges which we represent by 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 a node a is its own ancestor or descendant meaning that 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 ((Pearl, 1988; Cowell et al., 1999)). Furthermore, these Markov properties are equivalent to a *factorisation* of the joint distribution. By this we mean that if $X_1, \ldots, X_K$ are the variables represented by the nodes of the graph and $\mathrm{pa}(i)$ denotes the set of parents of X_i in the graph, the above Markov properties hold if and only if

$$P(X_1, \ldots, X_K) = \prod_{i=1}^{K} P(X_i | X_{\mathrm{pa}(i)}). \tag{1.8}$$

Some of the nodes can be decision variables, but these would typically not have any parents and would always be conditioned on. Also note that the requirement that the distribution of X given $F_X = x$ be degenerate at x ($x \neq \emptyset$) is not explicitly displayed in the graph and still has to be introduced as an implicit, externally specified, constraint (Dawid, 2002).

The graph in Figure 1.1 represents a situation where the assumptions (i) and (ii) of Section 3.2 are satisfied and the ACE can be identified by adjusting for C. We can see that $C \perp\!\!\!\perp F_X$ because they are non–descendants of each other and have no parents and $Y \perp\!\!\!\perp F_X | (X, C)$ because F_X is a non–descendant of Y and (X, C) are the parents of Y. The graph induces the following factorisation

$$P(Y, X, C | F_X) \quad = \quad P(Y|X, C) P(X|C, F_X) P(C).$$

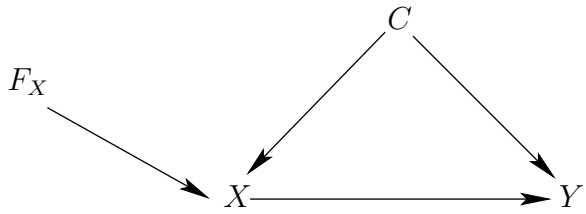

Figure 1.1. A situation where the set C is sufficient for identification of the ACE.

Depending on what value F_X takes, the distribution $P(X|C, F_X)$ is either observational or a one–point distribution on the value x. In order to read off the other conditional independencies that are implicit in the factorisation, we can use either the *moralisation* criterion (Lauritzen et al., 1990; Cowell et al., 1999) or alternatively, the method of d–separation (Pearl, 1988; Verma and Pearl, 1990). The former constructs an undirected (moral) graph $\mathcal{G}^m$ such that conditional independencies correspond to path separation in this undirected graph. The moral graph is constructed by adding an undirected edge between any two unconnected parents of a common child and removing the directions from all remaining edges. Any separation in the resulting undirected graph corresponds to a conditional independence in the underlying probabilistic model, where we say that "C separates A and B" if every (undirected) path between A and B contains nodes in C. These could equivalently be derived from the factorisation (1.8) but the graphical manipulations are often easier to carry out than the algebraic ones.

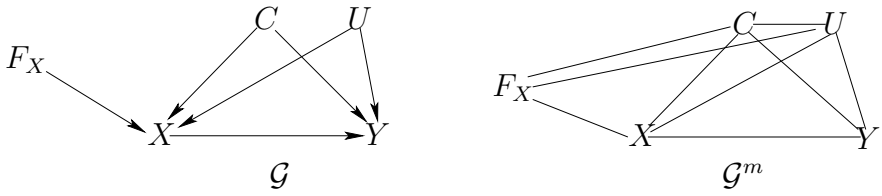

Figure 1.2. A graph $\mathcal{G}$ and corresponding moral graph $\mathcal{G}^m$ where C is not sufficient to identify the causal effect.

Consider the example in Figure 1.2. In such a situation, C is not sufficient to identify the causal effect as can be seen from the moral graph on the right: Y and F_X are not separated by $\{X, C\}$ because there is a path Y—

U—F_X linking them. Hence the condition $Y \perp\!\!\!\perp F_X|(X, C)$ is not satisfied. However, the conditions $Y \perp\!\!\!\perp F_X|(X, C, U)$ and $(C, U) \perp\!\!\!\perp F_X$ hold implying that if U were observable, an adjustment with regard to both (U, C) would be possible and yield valid causal inference.

Figure 1.3, on the other hand, depicts a situation where each of C or D alone is sufficient for adjustment. Here we have that $C \perp\!\!\!\perp F_X$ and $D \perp\!\!\!\perp F_X$, as the two variables in each statement are non–descendants of each other, and both $Y \perp\!\!\!\perp F_X|(X, C)$ and $Y \perp\!\!\!\perp F_X|(X, D)$, as can be seen from the corresponding moral graph $\mathcal{G}^m$. This situation illustrates the claim by Dawid (2002) that we do not need to adjust for 'all confounders' in the following sense. According to standard definitions of confounding that do not use a formal causal framework, each of C and D in Figure 1.3 are confounders in that they are not 'on the causal pathway' and they are both 'affecting' X and predicting Y. However, as we have shown, it is sufficient to adjust for only one of them in order to estimate the ACE..

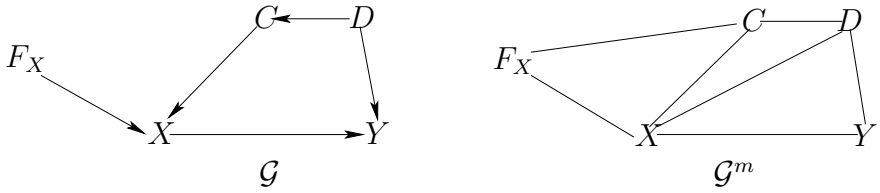

Figure 1.3. A graph $\mathcal{G}$ and corresponding moral graph $\mathcal{G}^m$ indicating that C or D are each on their own sufficient for adjustment.

5 Mendelian Randomisation

'Mendelian randomisation' denotes the random assortment of genes from parents to offspring that occurs during gamete formation and conception. This 'randomness' can be exploited to test for, or estimate, the causal effect of an intermediate phenotype that has a genetic component on a disease in situations where confounding between the phenotype and the disease status is believed to be likely and is not fully understood (Davey Smith and Ebrahim, 2003; Katan, 2004; Thomas and Conti, 2004). There are other uses of the method including the provision of information about alternative biological pathways to a disease (Davey Smith and Ebrahim, 2004; Davey Smith et al., 2005). However, our focus here will be on its use to test for and estimate the causal effect of an intermediate phenotype on a disease in order to inform public health interventions.

5.1 The Basic Idea

The notion of Mendelian randomisation that we will use derives from an idea put forth by Katan (1986). In the mid-1980s, there was much debate over the direction of an observed association between low serum cholesterol levels and cancer. The hypothesis of interest was that low serum cholesterol increases the risk of cancer but it is also plausible that hidden tumours lower cholesterol in future cancer patients or other lifestyle factors affect both cholesterol levels and cancer risk. (Katan, 2004). Katan noted that people with the rare genetic disease abetalipoproteinaemia, resulting in extremely low serum cholesterol levels, do not seem especially predisposed to getting cancer prematurely. It was known that the apolipoprotein E (APOE) gene is associated with cholesterol levels and that the E2 allele relates to lower levels than either E3 or E4. Crucially, by Mendel's Second Law (the law of assortment), E2 carriers should be no different from other genotypes in socioeconomic position, lifestyle and all other respects (this can be violated for various reasons and should always be checked in the light of background knowledge). Katan reasoned that 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.

The causality of the low cholesterol-cancer association was disproved by the subsequent large statin trials primarily concerned with the effects of *high* cholesterol levels on CHD risk (Scandinavian Simvastin Survival Study (4S), 1994; Heart Protection Study Collaborative Group, 2002), but the idea has been applied several times since and is what is now understood as *Mendelian randomisation* in the epidemiological literature. Katan's original idea was centred around hypothesis testing to confirm or disprove causality but the method is also used to estimate the size of the effect of the phenotype on the disease (Minelli et al., 2004) and, indeed, to compare this estimate with that obtained from observational studies in order to assess the extent to which confounding has been accounted for. Essentially, this approach exploits the idea that a genotype affecting the phenotype of interest, and thus indirectly affecting the disease status, is assigned randomly at meiosis, given the parents' genotype, independently of any possible confounding factors. It is well known in the econometrics and causal literatures (Bowden and Turkington, 1984) that these properties define an *instrumental* variable but additional fairly strong assumptions are required for unique identification of the causal effect of the phenotype on the disease status. These additional assumptions can take the form of *linearity* and *additivity* assumptions for all dependencies, as are typically assumed in econometrics applications but could also be assumptions about the *compliance* behaviour of subjects under study, as are

often made in the context of randomised trials with incomplete compliance (Angrist et al., 1996),

5.2 Instrumental Variables

We will present the basic properties that characterise an instrumental variable in terms of conditional independence statements. These conditions have been given in many different forms, using intervention indicators (Dawid, 2003) or counterfactuals (Greenland, 2000; Angrist et al., 1996; Robins, 1997) or linear structural equations (Goldberger, 1972; Pearl, 2000) and a comparison of some of these can be found in Galles and Pearl (1998). The conditions we give below are common to most instrumental variable methods but on their own they do not necessarily allow for identification of the ACE as we will discuss more fully in the following sections. For now, we will focus on these core assumptions and illustrate their meaning.

Core Conditions

Let X and Y be defined as above with the causal effect of X (e.g. homocysteine level) on Y (e.g. CHD) being of primary interest and F_X being the intervention indicator. Furthermore, let G be the variable that we want to use as the instrument (the MTHFR genotype in our case).

The following 'core conditions' that G has to satisfy (e.g. Greenland (2000); Dawid (2003)) assume the existence of a concrete, although possibly unobservable, variable (or set of variables) U[3] such that, under the observational regime i.e. under $F_X = \emptyset$,

1. $G \perp\!\!\!\perp U$, i.e. G must be (marginally) independent of U;

2. $G \not\perp\!\!\!\perp X$, i.e. G must not be (marginally) independent of X; and

3. $Y \perp\!\!\!\perp G \mid (X, U)$, i.e. conditionally on X and U, the instrument and the response are independent.

These alone do not allow us to infer anything about the intervention situation, i.e. about quantities under $F_X = x$. Hence, we need to supplement them with suitable further assumptions. We note that such extra assumptions are only implicit in the counterfactual approach of Greenland (2000) and also that of Pearl (2000). To motivate these additional assumptions, note that the conditional independencies of condition 1 and 3 are equivalent

[3]Note that Dawid (2003) points out that "there is no compelling reason to posit the existence of such an unobserved variable. To make this assumption is to say something non–trivial about how the world is. And even if it can be assumed to exist, there is no reason why the variable U should be essentially unique."

to the factorisation

$$P(Y, X, U, G|F_X = \emptyset) = \\ P(Y|X, U, F_X = \emptyset)P(X|U, G, F_X = \emptyset)P(U|F_X = \emptyset)P(G|F_X = \emptyset).$$

We now assume that if we change to the interventional setting, only the factor $P(X|U, G, F_X = \emptyset)$ changes to $P(X = x'|U, G, F_X = x) = I\{x' = x\}$ while the other terms remain the same. This reflects the crucial assumption that an intervention in X is possible without affecting the generation of the remaining variables in the system. Clearly, this would be quite idealistic in many circumstances. More formally it means that our intervention is such that

$$(G, U) \perp\!\!\!\perp F_X \text{ and } Y \perp\!\!\!\perp F_X|(X, U) \tag{1.9}$$

i.e. G and U are 'not on the causal pathway' from X to Y, and U would be sufficient for identifying the causal effect by adjustment if it were observable. As it is not we might call it a sufficient concomitant (Dawid, 2003). The distribution under intervention in X hence satisfies the following factorisation,

$$P(Y, X = x, U, G|F_X = x) = \\ P(Y|X = x, U, F_X = \emptyset)P(U|F_X = \emptyset)P(G|F_X = \emptyset),$$

where $P(Y, X = x', U, G|F_X = x) = 0$ if $x' \neq x$.[4]

Just as in the case of the assumptions required for confounder adjustment (Section 3.2), these conditions essentially have to be justified by subject matter background knowledge. Conditions 1 – 3 could be tested if U were observed, but otherwise they do not imply any testable independencies regarding the instrument G. In particular, they do not imply that G and Y are independent either marginally or conditionally on X alone (as has been assumed by Thomas and Conti (2004); Thompson et al. (2003) and implied by Foster (1997)). Moreover, conditions (1.9) must also be justified by background knowledge and will depend on what kind of actual intervention is being contemplated. Of course, U can be empty indicating that there is no need to adjust for confounding and hence no need to use an instrumental variable if X and Y can be simultaneously observed.

The typical Mendelian randomisation setting, where G corresponds to the genotype for phenotype X usually provides very detailed biological background knowledge to verify conditions 1 – 3. We know that genes are

[4] One could tentatively formulate alternative conditions to 1 – 3 avoiding the assumption of the existence of such a U as: 1A. $G \perp\!\!\!\perp F_X$, 2A. $G \not\perp\!\!\!\perp X$, 3A. $Y \perp\!\!\!\perp G|F_X = x$. This has not been considered yet, but see Pearl (2000, p.248).

randomly assigned at meiosis and can therefore be reasonably assumed not to be associated with anything concerning lifestyle factors confounding the relationship between X and Y, thus supporting condition 1. Condition 2 is typically verified by the fact that we only use genes with well-understood biological function in these applications and likewise, the biology can also help to exclude the possibility that G is affected by X (reverse causation), although the direction of this association is not crucial. Likewise, the existence of alternative biological pathways between G and Y other than through X can often be ruled out thus supporting condition 3. The assumption that we can intervene in the phenotype, setting it to a fixed value so that none of the other variables are affected, is generally more problematic in terms of justification.

Graphical Representation

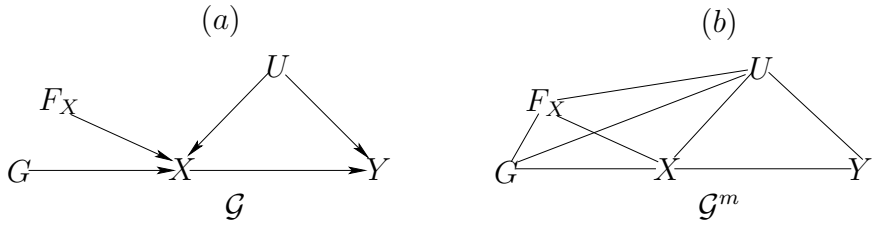

Figure 1.4. A graph $\mathcal{G}$ and corresponding moral graph $\mathcal{G}^m$ indicating that the core conditions for G to be an instrument are satisfied.

Figure 1.4 (a) shows a DAG involving G, X, Y and U that satisfies the core conditions 1 – 3, where the moral graph (b) in particular shows that condition 3 holds as Y and G are separated by (X, U) despite the moral edges that have to be added. In addition, by including the node F_X in the way shown, we ensure (1.9). The conditional independence restrictions imposed by the graph in Figure 1.4 (a) are equivalent to a factorisation of the joint density in the following way:

$$P(Y, X, U, G|F_X) = P(Y|U, X)P(X|U, G, F_X)P(U)P(G). \qquad (1.10)$$

From this, or from the moral graph in Figure 1.4 (b), it can be seen (by integrating out Y and conditioning on X) that $G \not\perp\!\!\!\perp U|(X, F_X = \emptyset)$, for instance. Similarly, by integrating out X and conditioning on Y, we have that $G \not\perp\!\!\!\perp U|(Y, F_X = \emptyset)$ despite $P(G, U) = P(G)P(U)$ (under either, $F_X = x$ or $F_X = \emptyset$). 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. Hence the additional (moral) edge between G and U in the moral graph.

5.3 Linear No–Interaction Models

As mentioned above, the core conditions of Section 5.2 alone are not sufficient to allow us to obtain the causal effect in the presence of confounding for which we cannot adjust. More assumptions have to be made, typically with regard to the parametric shape of the relationships amongst the variables. In this section we continue with the simplistic situation described in Section 3.3 where a potential confounder is added, but this time we call it U to emphasise that it is unobservable.

As in Section 3.3 (with U replacing C), our model assumptions are that $U \perp\!\!\!\perp F_X$, i.e. we assume (1.9), which as explained earlier (cf. equation (1.5)) can be regarded as stating that U is not on the "causal pathway" from X to Y, and

$$\begin{aligned} E(Y|X=x,U=u,F_X=\emptyset) &= \alpha+\beta_1 x+\beta_2 u \quad \text{and} \\ E(X|G=g,U=u,F_X=\emptyset) &= \gamma+\delta_1 g+\delta_2 u. \end{aligned}$$

In addition we assume that the first expectation is the same if we intervene in X i.e.

$$E(Y|U=u,F_X=x)=\alpha+\beta_1 x+\beta_2 u.$$

reflecting assumption (1.9), i.e. that $Y \perp\!\!\!\perp F_X|(X,U)$. As before, β_1 is the causal parameter of interest here.

As we cannot adjust for U and a regression of Y on X alone does not yield the correct parameter (as shown in Section 3.3), we instead consider a regression of Y on G alone based on observational data, i.e. under $F_X=\emptyset$ (omitted from the conditioning sets below for brevity). This corresponds to

$$\begin{aligned} E(Y|G=g) &= E_{(X,U)|G=g}E(Y|X,U,G=g) \\ &= E_{U|G=g}E_{X|U,G=g}E(Y|X,U) \quad \text{since} \quad Y \perp\!\!\!\perp G|(X,U) \text{ by 3} \\ &= E_U E_{X|U,G=g}E(Y|X,U) \quad \text{since} \quad U \perp\!\!\!\perp G \text{ by 1} \\ &= E_U(\alpha+\beta_1(\gamma+\delta_1 g+\delta_2 U)+\beta_2 U) \\ &= \alpha+\beta_1\gamma+\beta_1\delta_1 g+(\beta_1\delta_2+\beta_2)\mu_U \\ &= \alpha^*+\beta_1\delta_1 g. \end{aligned}$$

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 based on observational data 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 the causal parameter of interest, β_1, can be estimated consistently from the ratio of these two regression coefficients where the requirement $\delta_1 \neq 0$ is ensured by core condition 2. Note that the previous argument and model assumptions can easily be generalised to the case where X and G are binary. The case where Y is binary is more difficult and addressed below.

We should emphasise that, given that we are using the IV approach in the first place because confounding is not sufficiently understood, it seems unrealistic to believe that one would be willing to make such strong assumptions about U as are required for the above, in particular with regard to the parametric shape of the dependence of Y on U. Note that generalisations to the non–linear case have been developed in the econometrics literature but cannot necessarily be used for the present purpose as they are often targeted at situations with measurement errors (Amemiya, 1974; Hansen and Singleton, 1982).

5.4 More Realistic Situations

The instrumental variable approach for linear models without interactions as described above is well known from econometric theory. However, it is only of limited value for typical epidemiological applications where the primary aim of an investigation into the causal effect of a risk factor on a disease is to inform public health interventions. Firstly, the response Y is often a binary variable. Secondly, the data often arise from case–control studies with retrospective sampling (i.e. conditional on Y). Based on the above framework for causal inference we have shown (Didelez and Sheehan, 2005) that:

- If the core conditions 1 – 3 are satisfied, a test for *no causal effect* of the intermediate phenotype on the disease can be performed by testing that $Y \perp\!\!\!\perp G$, regardless of whether the data have been collected retrospectively and regardless of how the relevant variables have been measured. This was in fact the original idea of Katan (1986).

- If G, X and Y are all binary, the well-known ratio estimate derived in Section 5.3 cannot be applied. In fact, it is not straightforward

to even specify the causal parameter in the latter case and the case-control scenario is further complicated by the fact that only odds ratios can be used. This has often been overlooked in the epidemiological literature.

Instead, bounds for the ACE can be derived as in Robins (1989); Manski (1990); Balke and Pearl (1994); Lauritzen (2000) and without assumptions about counterfactuals as in Dawid (2003). These can be modified to account for the case–control situation when gene frequencies $P(G)$ are available (Didelez and Sheehan, 2005) and can also be used as a rough test to rule out poor instruments. (Pearl, 2000) The calculations become computationally expensive when some variables have more than two categories and are intractable for continuous variables.

- An approximate check for confounding can be carried out when Y is binary and X continuous as in many epidemiological applications.

5.5 Problems with Mendelian Randomisation

The limitations of Mendelian randomisation, 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 (Davey Smith and Ebrahim, 2003, 2004; Thomas and Conti, 2004; Davey Smith et al., 2005; Nitsch et al., 2006). However, existing approaches to testing and estimating the causal effect have not been formally challenged. Background knowledge is always required to verify untestable assumptions in order to make causal inferences. Mendelian randomisation applications have an advantage in that substantial biological background information can frequently be exploited in order to check that a particular genotype satisfies the conditions for an instrumental variable. It is unlikely that our simple model of Figure 1.4 will pertain, in practice. The common complex diseases that are of most interest from a public health perspective are generally multifactorial in nature and the definition of disease outcome itself is often ambiguous. As before, we can use directed acyclic graphs to represent the conditional dependencies that we believe are implied by the underlying biology and check the core conditions visually. We will illustrate this with a few examples below. Note however that, even if the core conditions would appear to be satisfied, the additional parametric assumptions that permit estimation of the causal effect of interest may not be justifiable.

Linkage disequilibrium refers to an association between alleles at different loci across the population and can be due either to tight *linkage* (i.e.

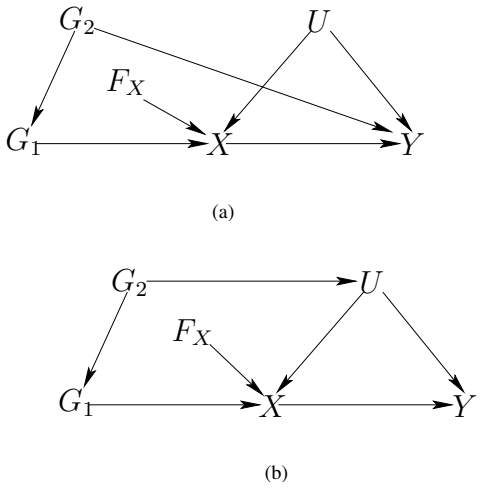

Figure 1.5. The instrument G_1 is in linkage disequilibrium with G_2 influencing Y, directly as in (a), or influences Y indirectly via the confounder U, as in (b).

because the loci are physically close on the chromosome and thus tend to be inherited together) or to other reasons such as natural selection, assortative mating, and migration, for example Lynch and Walsh (1998). If our chosen instrument G_1 is in linkage disequilibrium with another gene G_2 which has a *direct* influence on the disease Y, condition 3, $Y \perp\!\!\!\perp G_1 | (X, U)$, might be violated as shown in Figure 1.5 (a). If G_2 affects Y indirectly via a route other than through X, (Figure 1.5 (b)), condition 1, $G \perp\!\!\!\perp U$ might be violated. Note that even if the conditions appear to hold, linkage disequilibrium can cause attenuation of the genotype–phenotype association leading to poor inference.

Pleiotropy is the phenomenon whereby a single gene may influence several traits. If the chosen instrument G is associated with another intermediate phenotype which is also associated with the disease Y (Figure 1.6 (a)), condition 3 $Y \perp\!\!\!\perp G | (X_1, U)$, is again violated. As before, the association of X_2 with Y can also be via U (e.g. pleiotropic effects might influence consumption of tobacco or alcohol, for example) to violate condition 1 as in Figure 1.6 (b).

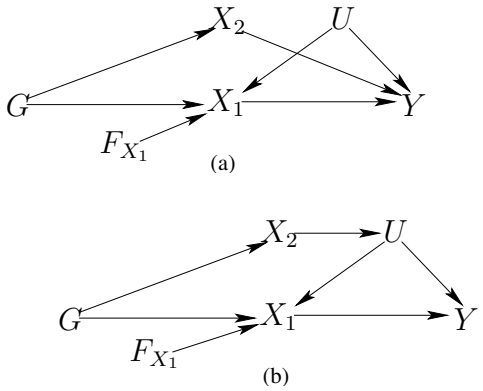

Figure 1.6. G has pleiotropic effects X_1 and X_2 where (a) both have a direct effect on the outcome Y of interest, or (b) X_1 has a direct effect but X_2 has an indirect effect via the confounder U.

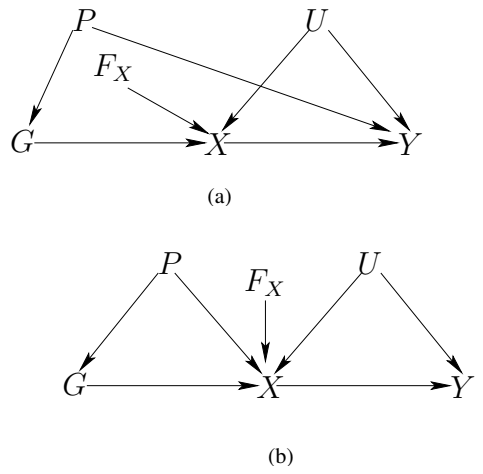

Figure 1.7. Two examples of population stratification where one of the conditions for G to be an instrument is violated (a) and all conditions are satisfied (b).

Population Stratification, referring to the co–existence of different disease rates and allele frequencies within subgroups of individuals, could lead to an association between the two at the population level which in turn can result in confounding of the genotype–disease association. That the disease rates and allele frequencies are different for different population strata is represented in Figure 1.7 (a) by the arrows from P into G as well as into Y. We see from this graph 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 1.7 (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, as a result.

6 Summary and Discussion

The aim of this article is to justify our opinion that epidemiology in general, and the applications of Mendelian randomisation in particular, can benefit greatly from a formal framework for causal inference. We have presented three possible approaches. The first two, the $do(\cdot)$–operator and the intervention indicator F_X are very similar, with F_X being more general and explicit. The third, based on counterfactuals, is philosophically quite different especially with respect to the type of assumptions required. We clearly lean towards the first two. Note that the counterfactual language has been used very successfully to solve many intricate questions of causal inference, especially in the field of epidemiology, although it is still not widely adopted. Another formal approach that we have not considered here is provided by functional models (see Pearl, 2000, chapters 5–10) which include linear structural equation models (Haavelmo, 1943; Goldberger, 1972) as a special case. These are closely related to counterfactual reasoning and discussed in Dawid (2002), for example. Note that the models we considered in Sections 3.3 and 5.3 can be regarded as (partial) sturctural mean models in the sense that they provide equations for the mean of the response variable that remain stable under interventions in X, but this alone does not permit construction of the counterfactual responses for individuals.

We strongly advocate the adoption of a formal approach to causal inference in epidemiological applications—whichever of the above is favoured. Such a framework allows explicit specification of what the causal aim under investigation actually is: this cannot usually be expressed in a purely probabilistic way with 'normal' conditional probabilities. In the typical Mendelian randomisation setting, the causal aim would be some public health intervention and thus it would seem particularly desirable to make such inter-

ventions an integral part of the statistical model via decision variables such as F_X. For one thing, such an explicit representation can aid communication between the biologist and the statistician. Moreover, a formal language of causality allows explicit formulation of the conditions under which the target of inference can be attained. This, again, seems important with regard to facilitating the discussions and investigations that statisticians and subject matter experts have to carry out *together* when assessing whether the necessary conditions are satisfied and hence whether the ensuing causal conclusions are valid. Recall, for example, that our core condition 3 has been misinterpreted in the literature due to reliance on imprecise verbal descriptions. A particular advantage of the F_X indicator is that we can express the assumption that the core conditions 1 – 3 hold under $F_X = \emptyset$ and that *in addition* to these core conditions, (1.9) has to hold in order to specify which variables remain unaffected by the particular intervention that is being contemplated. The latter condition is implicit in Pearl's $do(\cdot)$ formalism and in the counterfactual approach.

We argue and confirm that Mendelian randomisation can often be reasonably assumed to satisfy the instrumental variable conditions. Moreover, subject matter background knowledge can be encoded using directed acyclic graphs to facilitate verification of the core conditions by visual inspection, and violations such as those of Section 5.5 can easily be identified. But there is also reason for some concern as to whether the public health interventions that can be carried out in practice satisfy conditions such as (1.9). For example, we cannot expect to be able to *fix* homocysteine at a prespecified level for the whole population. We can fix the amount of folic acid added to flour, but the amount of bread people eat and the amount by which folic acid affects the homocysteine level of individuals will vary. Such assumptions hence have to be scrutinised within the context of a specific application and with a concrete intervention in mind. Further research is required to address extending the notion of interventions in this respect.

On a more technical level, we showed in Didelez and Sheehan (2005) that under the above conditions, testing for a phenotype-disease causal effect by testing for a genotype-disease association is reasonable in all cases for practical purposes and that for calculation of the average causal effect, one must rely on additional strong parametric assumptions such as linearity and no interactions. In Sections 3.3 and 5.3 of the present article we have considered the linear case in some more detail in order to exemplify the use of the intervention indicator within this simple and well–known model class. But we emphasise that this model class is typically not useful for Mendelian randomisation settings, where non–linear models are called for and interactions may play a role. We then have to deal with even more

technical problems that have yet to be formally tackled, For example, 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. "There is, in fact, no agreed upon generalisation of instrumental variables to non-linear systems" (see Pearl (2000), p.248). However, the technical issues cannot be satisfactorily addressed if the causal questions are not clearly stated in the first place. We argue that this cannot be achieved consistently without adopting some formal framework for causality.

Acknowledgments

The authors acknowledge research support from the Centre for Advanced Study of the Norwegian Academy of Science and Letters.

Vanessa Didelez
Department of Statistical Sciences, University College London, WC1E 6BT, UK.
vanessa@stats.ucl.ac.uk

Nuala A. Sheehan
Departments of Health Sciences and Genetics, University of Leicester, Leics LE1 6TP, UK.
nas11@leicester.ac.uk

BIBLIOGRAPHY

Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group (1994). 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.

Amemiya, T. (1974). The nonlinear two-stage least-squares estimator. *Journal of Econometrics*, 2:105–110.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Balke, A. A. and Pearl, J. (1994). Counterfactual probabilities: Computational methods, bounds and applications. In Mantaras, R. and Poole, D., editors, *Proceedings of the 10th Conference on Uncertainty in Artificial Inteligence*, pages 46–54.

Bowden, R. and Turkington, D. (1984). *Instrumental Variables*. Cambridge University Press, Cambridge, U.K.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. Springer-Verlag, New York, Inc.

Czeizel, A. E. and Dudás, I. (1992). Prevention of the first occurence of neural-tube defects by periconceptional vitamin supplementation. *New England Journal of Medicine*, 327:1832–35.

Davey Smith, G. and Ebrahim, S. (2003). Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, 32:1–22.

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

Davey Smith, G., Ebrahim, S., Lewis, S., Hansell, A., Palmer, L., and Burton, P. (2005). Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet*, 366:1484–1498. (a).

Dawid, A. P. (1979). Conditional independence in statistical theory (with Discussion). *Journal of the Royal Statistical Society, Series B*, 41:1–31.

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95:407–448.

Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review*, 70:161–189.

Dawid, A. P. (2003). Causal inference using influence diagrams: the problem of partial compliance. In Green, P. J., Hjort, N. L., and Richardson, S., editors, *Highly Structured Stochastic Systems*, pages 45–81. Oxford University Press, Oxford, UK.

Dawid, A. P. and Didelez, V. (2005). Identifying the consequences of dynamic treatment strategies. Research Report 262, Department of Statistical Sciences, University College London.

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

Fisher, R. (1926). *The Design of Experiments*. Oliver & Boyd, Edinburgh, 1st edition.

Ford, E., Smith, S., Stroup, D., Steinberg, K., Mueller, P., and Thacker, S. (2002). Homocysteine and cardiovascular disease: a systematic review of the evidence with special emphasis on case-control studies and nested case-control studies. *International Journal of Epidemiology*, 31:59–70.

Foster, E. (1997). Instrumental variables for logistic regression: An illustration. *Social Science Research*, 26:487–504.

Galles, D. and Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3:151–182.

Goldberger, A. (1972). Structural equation methods in the social sciences. *Econometrica*, 40:979–1001.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, 29:722–729.

Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, 10:37–48.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11:1–12.

Hansen, L. and Singleton, R. (1982). Generalized instrumental variable estimation of non-linear rational expectation models. *Econometrica*, 50:1269–1286.

Heart Protection Study Collaborative Group (2002). MRC/BHF heart protection study of cholesterol lowering with simvastin in 20,536 high-risk individuals. *Lancet*, 360:7–22.

Hernán, M. (2004). A definition of causal effect for epidemiologic research. *Journal of Epidemiology and Community Health*, 58:265–271.

Hernán, M., Hernández-Diaz, S., and Robins, J. (2004). A structural approach to selection bias. *Epidemiology*, 15:615–625.

Homocysteine Lowering Trialists' Collaboration (1998). Lowering blood homocysteine with folic acid based supplements: meta-analysis of randomized controlled trials. *BMJ*, 316:894–898.

Katan, M. B. (1986). Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet*, i:507–508.

Katan, M. B. (2004). Commentary: Mendelian randomization, 18 years on. *International Journal of Epidemiology*, 33:10–11.

Lauritzen, S., Dawid, A., Larsen, B., and Leimer, H. (1990). Independence properties of directed Markov fields. *Networks*, 20:491–505.

Lauritzen, S. L. (2000). Causal inference from graphical models. In Barndorff-Nielsen, O. E., Cox, D. R., and Kluppelberg, C., editors, *Complex Stochastic Systems*, chapter 2, pages 63–107. Chapman & Hall.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Inc., USA.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings*, 80:319–323.

Minelli, C., Thompson, J., Tobin, M., and Abrams, K. (2004). An integrated approach to the Meta-Analysis of genetic association studies using Mendelian randomization. *American Journal of Epidemiology*, 160:445–452.

MRC Vitamin Study Research Group (1991). Prevention of neural tube defects: Results of the medical research council vitamin study. *Lancet*, 338:131–137.

Neyman, J. (1923). Sur les applications de la thar des probabilités aux experiences agariales: essay des principles. *Statistical Science*, 5:463–472. 1990. Excerpts reprinted in English. (D.Dabrowska and T.Speed, Trans.).

Nitsch, D., Molokhia, M., Smeeth, L., DeStavola, B. L., Whittaker, J. C., and Leon, D. A. (2006). Limits to causal inference based on Mendelian randomization: a comparison with randomised controlled trials. *American Journal of Epidemiology*, 163:397–403.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA.

Pearl, J. (1993). Graphical models, causality and intervention. *Statistical Science*, 8:266–269.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82:669–710.

Pearl, J. (2000). *Causality*. Cambridge University Press.

Peto, R., Doll, R., Buckley, J., and Sporn, M. (1981). Can dietary beta-carotene materially reduce human cancer rates? *Nature*, 290:201–208.

Robins, J. (1989). The analysis of randomized and nonrandomized aids treatment trials using a new approach to causal inference in longitudinal studies. In Sechrest, L., Freeman, H., and Mulley, A., editors, *Health Service Research Methodology. A Focus on AIDS*, pages 113–159. U.S. Public Health Service, Washington, D.C.

Robins, J. (1997). Causal inference from complex longitudinal data. In Berkane, M., editor, *Latent Variable Modeling with Applications to Causality*, pages 69–117. Springer-Verlag, New York.

Robins, J. (2001). Data, design and background knowledge in etiologic inference. *Epidemiology*, 11:313–320.

Rothman, K. and Greenland, S. (1998). *Modern Epidemiology*. Lippincott-Raven, Philadelphia, 2nd edition.

Rubin, D. (1974). Estimating causal effexts of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.

Rubin, D. (1978). Bayesian inference for causal effects:the role of randomization. *Annals of Statistics*, 6:34–58.

Scandinavian Simvastin Survival Study (4S) (1994). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease. *Lancet*, 344:1383–1389.

Scholl, T. O. and Johnson, W. G. (2000). Folic acid: influence on the outcome of pregnancy. *American Journal of Clinical Nutrition*, 71 (Suppl.):12955–3035.

Stanley, F., Blair, E., and Alberman, E. (2000). *Cerebral Palsies: Epidemiology and Causal Pathways*. Mac Keith Press, London, U.K.

Thomas, D. and Conti, D. (2004). Commentary: The concept of "Mendelian randomization". *International Journal of Epidemiology*, 33:21–25.

Thompson, J., Tobin, M., and Minelli, C. (2003). 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, (http://www.prw.le.ac.uk/research/HCG/getechrep.html).

Tobin, M., Minelli, C., Burton, P., and Thompson, J. (2004). Commentary: Development of Mendelian randomization: from hypothesis test to "Mendelian deconfounding". *International Journal of Epidemiology*, 33:26–29.

Verma, T. and Pearl, J. (1990). Causal networks: semantics and expressiveness. In Bonisonne, P., Henrion, M., Kanal, L., and Lemmer, J., editors, *Uncertainty in Artificial Interrligence 6*, pages 255–268. North-Holland, Amsterdam.

Weinberg, C. (1993). Towards a clearer definition of confounding. *American Journal of Epidemiology*, 137:1–8.