Open Access

Stephen Burgess* and Dylan S. Small

Predicting the Direction of Causal Effect Based on an Instrumental Variable Analysis: A Cautionary Tale

DOI 10.1515/jci-2015-0024

Abstract: An instrumental variable can be used to test the causal null hypothesis that an exposure has no causal effect on the outcome, by assessing the association between the instrumental variable and the outcome. Under additional assumptions, an instrumental variable can be used to estimate the magnitude of causal effect of the exposure on the outcome. In this paper, we investigate whether these additional assumptions are necessary in order to predict the direction of the causal effect, based on the direction of association between the instrumental variable and the outcome, or equivalently based on the standard (Wald) instrumental variable estimate. We demonstrate by counterexample that if these additional assumptions (such as monotonicity of the instrument–exposure association) are not satisfied, then the instrumental variable–outcome association can be in the opposite direction to the causal effect for all individuals in the population. Although such scenarios are unlikely, in most cases, a definite conclusion about the direction of causal effect requires similar assumptions to those required to estimate a causal effect.

Keywords: instrumental variables, Mendelian randomization, Simpson's paradox
Classification: causal inference

1 Introduction

Instrumental variable analysis is a technique for obtaining causal inferences about the relationship between a putative causal risk factor (referred to as an exposure) and an outcome. An advantage of instrumental variable analysis is that it does not require the specification of a parametric model in order to make a causal claim: an association between an instrumental variable and the outcome implies a causal effect of the exposure on the outcome [1, 2]. However, in order to estimate a causal effect of the exposure on the outcome, further assumptions are required depending on the causal parameter that is targeted. In this paper, we ask the question: if one does not want to estimate a causal effect parameter, but only to conclude the direction of the causal effect (that is, does increasing the exposure lead to increases or decreases in the outcome?), what assumptions are required? In particular, can the directions of association between the instrumental variable and the exposure, and between the instrumental variable and the outcome, be used to predict the direction of causal effect of the exposure on the outcome?

This question is particularly relevant for Mendelian randomization, the use of genetic variants are instrumental variables [3, 4]. Several authors have advocated reporting the presence or absence of an association between the genetic variant(s) and the outcome as the primary analysis result, rather than a causal effect estimate [5, 6]. This is analogous to performing an intention-to-treat analysis in a randomized trial [7]. The motivation for this is that the claim of a causal effect requires fewer assumptions than the estimation of a causal effect [8, 9], and the magnitude of the causal estimate is of secondary importance, as the quantitative effect of intervening on the exposure in practice is likely to differ from the causal estimand of the instrumental variable analysis [10]. For example, the effect of reducing low-density lipoprotein cholesterol

***Corresponding author: Stephen Burgess,** Department of Public Health and Primary Care, University of Cambridge, 2 Worts Causeway, Cambridge, Cambridge CB1 8RN, United Kingdom of Great Britain and Northern Ireland,
E-mail: sb452@medschl.cam.ac.uk
http://orcid.org/0000-0001-5365-8760
Dylan S. Small, University of Pennsylvania, Pennsylvania, PA, USA

(LDL-c) on coronary heart disease (CHD) risk by taking statin drugs depends on the choice of statin, the dosage (amount and frequency), the duration of treatment, the patient group, and so on [11]. The quantitative estimate from a meta-analysis of the effect of statin treatment for five years or more duration in a primary prevention context is that a 30% reduction in LDL-c leads to a 27% (95% confidence interval, 23 to 30%) relative reduction in CHD risk [12]. In contrast, the Mendelian randomization estimates based on one of five genetic variants scaled to a 30% reduction in LDL-c range from a 55% to a 73% relative reduction in CHD risk [10], with the corresponding estimate based on all five genetic variants being a 67% (95% confidence interval, 54 to 76%) relative reduction in CHD risk [13]. There are several reasons for differences between the estimates: the Mendelian randomization estimates rely on between eight- and twenty-fold extrapolations of the genetic associations with disease risk, and the Mendelian randomization estimate represents a life-long reduction in LDL-c concentrations. As CHD often results from a long-term build-up of fatty deposits in the coronary arteries, it is not surprising that an estimate corresponding to life-long intervention in LDL-c concentrations is greater than an estimate corresponding to a more limited intervention.

The direction of the causal effect is informative as to whether interventions to increase or decrease the exposure should be prioritized; for a pharmaceutical intervention on an exposure that is a gene product (as if often the case in Mendelian randomization), this determines whether an inhibitor or a promoter of the genetic pathway is needed [14]. It would be valuable if the direction of the instrumental variable–outcome association was not only a test of the causal null hypothesis, but also predictive of the direction of the causal effect. However, as we demonstrate here, this is not always the case. In this paper, we outline scenarios in which the instrumental variable–outcome association (and equivalently the standard instrumental variable estimate) is in the opposite direction to the causal effect of the exposure on the outcome.

2 Definition of an instrumental variable

Two sets of assumptions have been proposed for defining an instrumental variable. We refer to these as the graphical assumptions and the counterfactual assumptions. We denote the instrumental variable as Z, the exposure as X, the outcome as Y, and confounders of the exposure–outcome relationship as U.

2.1 Graphical assumptions

The graphical assumptions (for example, see [15, 16]) require an instrumental variable to be:

(i) associated with the exposure: $Z \not\perp\!\!\!\perp X$;
(ii) independent of the confounders: $Z \perp\!\!\!\perp U$;
(iii) independent of the outcome conditional on the exposure and confounders: $Z \perp\!\!\!\perp Y \mid X, U$.

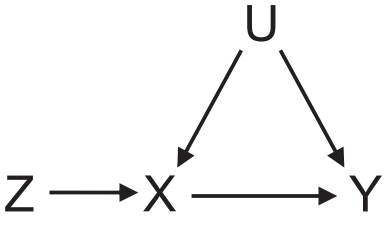

Figure 1: Directed acyclic graph of graphical instrumental variable assumptions.

A directed acyclic graph illustrating these assumptions is given as Figure 1. These assumptions imply that the joint distribution of Y, X, U, Z factorizes as:

$$p(y, x, u, z) = p(y|u, x)p(x|u, z)p(u)p(z). \quad (1)$$

In order to relate the observational distribution of these variables to the distribution under intervention in the exposure, the additional structural assumption has been proposed:

$$p(y,u,z,x|\mathrm{do}(X=x_0)) = p(y|u,x_0)1(X=x_0)p(u)p(z) \quad (2)$$

where $1(.)$ is the indicator function and the $\mathrm{do}(X=x_0)$ function represents intervention to set the value of the exposure to x_0. This assumption ensures that intervening on the exposure does not affect the distributions of any other variables except the conditional distribution of the outcome [17].

2.2 Counterfactual assumptions

Counterfactual values of the exposure and outcome are denoted as follows: $X(z)$ is the exposure when $Z=z$, and $Y(x,z)$ is the outcome when $X=x$ and $Z=z$. The counterfactual assumptions (see [2]) require an instrumental variable to be:

(i) associated with the exposure: $X(z)$ is a non-trivial function of z for at least some of the population;
(ii) independent from potential values of the exposure and outcome: $Z \perp\!\!\!\perp X(z), Z \perp\!\!\!\perp Y(x,z)$;
(iii) and to only influence the outcome via the exposure (the exclusion restriction assumption $Y(x,z) = Y(x)$;

where $Y(x)$ is the outcome when $X = x$. These assumptions imply that an instrumental variable cannot have a direct effect on the outcome, but instead any effect is mediated via the exposure [18].

2.3 Comparison of assumptions

The graphical assumptions require the explicit specification of the confounders, which is not required by the counterfactual assumptions. The graphical assumptions are expressed in terms of the observational distribution of the variables, meaning that an additional structural assumption is required to express a causal effect. The counter-factual assumptions are expressed in terms of counterfactual variables, and so there is a natural connection to causal effects under the consistency assumption that each variable takes its relevant counterfactual value under the observational regime.

An instrumental variable can satisfy the graphical assumptions without satisfying the counterfactual assumptions (for example, see [19] or Section 3.6). However, if an instrumental variable satisfies the counterfactual assumptions then it must satisfy the graphical assumptions. This can be seen by considering the contrapositive – if the graphical assumptions are violated, then the counterfactual assumptions are violated (if graphical assumption 2 is violated, then counterfactual assumption 2 is violated; and the same for assumptions 1 and 3).

A hypothetical (but realistic) example of a genetic variant that satisfies the graphical instrumental variable assumptions but not the counterfactual assumptions is as follows: the genetic variant affects the exposure, and additionally is partially correlated with another variant that also affects the exposure. Hence, there is no causal pathway from the original genetic variant (the proposed instrument) to the outcome other than that via the exposure. However, the genetic variant is correlated with the counterfactual values of the exposure on setting the value of the genetic variant (as this would not affect the value of the correlated variant).

2.4 Additional assumptions for the estimation of a causal effect

We assume throughout this paper that the instrumental variable Z is binary, taking values 0 and 1. This restriction is purely for clarity of presentation; the findings of this paper hold equally for non-binary instrumental variables. We assume that the exposure X is either binary (as is common in the use of

instrumental variables in randomized trials), or continuous (as is common in Mendelian randomization). The outcome Y is assumed to be continuous.

We define three further properties relevant to the estimation of a causal effect: the stable unit treatment value assumption (SUTVA), monotonicity, and homogeneity. The SUTVA is necessary to define a causal effect in a consistent way. Monotonicity and homogeneity are two additional assumptions; each of these enables the identification of a causal effect using an instrumental variable.

SUTVA: The stable unit treatment value assumption states that the potential outcomes for each individual should be unaffected by how the exposure was assigned, and unaffected by variables in the model relating to other individuals [20]. It is informally referred to as "no multiple versions of treatment," meaning that the effect on the outcome will be the same for all changes in the exposure no matter how the exposure is intervened on [21].

The individual causal effect of the exposure on the outcome for a binary exposure is $Y(1) - Y(0)$, where $Y(x)$ is the potential outcome when $X = x$. For a continuous exposure, the causal effect is $Y(x_1) - Y(x_0)$. The causal effect is linear if $Y(x + 1) - Y(x)$ is constant for all values of x, and monotone is $Y(x_1) - Y(x_0)$ is always positive (or always negative) for all $x_1 > x_0$. The SUTVA is a necessary assumption whenever a causal effect is estimated; without it, the causal effect is not well defined.

Monotonicity: Monotonicity is the property that the potential values of the exposure (these are counterfactual values, as for each individual only one value of the exposure can be observed) form an increasing function of the instrumental variable for all individuals in the population (or equivalently, a decreasing function for all individuals). If the exposure (X) and instrumental variable (Z) are both binary, then the population can be divided into four categories, known as principal strata [22]. These categories are called always-takers ($X = 1$ for both values of Z), never-takers ($X = 0$ for both values of Z), compliers ($X = 0$ when $Z = 0$ and $X = 1$ when $Z = 1$), and defiers ($X = 1$ when $Z = 0$ and $X = 0$ when $Z = 1$). In a randomized trial, the instrumental variable is typically random allocation to treatment, and the exposure is treatment received. Compliers are so-called as they "comply" with treatment assignment, in that exposure is present ($X = 1$) if they are randomized to exposure ($Z = 1$), and absent ($X = 0$) if they are randomized to no exposure ($Z = 0$). Defiers do exactly the opposite. The monotonicity assumption in this case is that there are no defiers in the population.

Homogeneity: Homogeneity refers to the similarity of the individual causal effect for different units in the population. A strong version of the homogeneity assumption is that the causal effect of the exposure on the outcome has the same magnitude in all individuals [8]. A weaker version is that there is no additive effect modification by the instrumental variable at different values of the exposure [23].

2.5 Instrumental variable estimate

The standard instrumental variable estimate with a binary instrumental variable (often called the Wald estimate [24]) is:

$$\frac{\widehat{\mathbb{E}}[Y|(Z=1)] - \widehat{\mathbb{E}}[Y|(Z=0)]}{\widehat{\mathbb{E}}[X|(Z=1)] - \widehat{\mathbb{E}}[X|(Z=0)]}.$$

where $\hat{\mathbb{E}}[.]$ represents an estimate of the expectation of the random variable. The numerator in the ratio estimate is the estimated association of the instrumental variable with the outcome; the denominator is the estimated association of the instrumental variable with the exposure.

Under the assumption of monotonicity (in the context of a randomized trial, if there are no defiers), then the instrumental variable estimate targets the average causal effect in the complier population (also known as the complier-averaged causal effect, or the local average treatment effect) [25]. For a binary exposure, this is the causal effect:

$$\mathbb{E}[Y(1) - Y(0) \mid \text{complier}], \tag{3}$$

where the expectation is taken across compliers.

Alternatively, under the assumption of homogeneity, the instrumental variable estimate targets an average causal effect in the population. In this case, the monotonicity assumption is not required. For a continuous exposure, if the effect of the exposure on the outcome is linear (that is, the marginal structural model for the outcome as a function of the exposure is linear), then the causal effect is the same at all levels of the exposure. The instrumental variable estimate targets then targets the average causal effect $\mathbb{E}[Y(x+1) - Y(x)]$.

3 Predicting the direction of the causal effect

We recall that the objective of this paper was to consider under what conditions the instrumental variable–exposure and instrumental variable–outcome associations predict the direction of the causal effect of the exposure on the outcome. We proceed to consider situations in which the instrumental variable estimate has the same or a different sign to that of the causal effect.

3.1 Non-monotone causal effect

If the causal effect of the the exposure on the outcome is non-monotone (for instance, it is positive for some individuals in the population, but negative for others; or else it is positive at some values of the exposure, but negative at other values), then there is no single `direction of causal effect'. The sign of the instrumental variable estimate will depend on the sample population, and may differ, for example, in a healthy population versus in a hospital-derived cohort. Hence, researchers should be cautious when extrapolating the results of their study to an external population. This issue is not unique or specific to instrumental variable analysis. One way of addressing this problem is to restrict the analysis to a subset of the population for which the causal effect would be expected to be monotone; however, this cannot be done by conditioning on the value of the exposure or outcome, as that may induce violations of the instrumental variable assumptions in the ascertained population.

Several exposures potentially have non-linear effects on outcomes. An example is the effect of body mass index on mortality – extreme low and high body mass index are both associated with increased mortality – although it is unclear to what extent the U-shaped relationship between body mass index and mortality is predicated by a causal effect of low body mass index on mortality, and how much this is reverse causation [26]. An exposure may have different directions of effect for different individuals in the population if there is an interaction with another variable. A plausible example is the effect of blood sugar levels on mortality: moderate to low blood sugar is likely to be beneficial for health outcomes for most of the population, but it may be harmful for diabetics.

For the remainder of the paper, we only consider monotone exposure–outcome relationships, so that the direction of causal effect is the same for all individuals in the population, and the objective of the paper is well-defined. We use the word "monotonicity" with respect to the instrumental variable–exposure relationship.

3.2 Homogeneous linear causal effect

If the causal effect of the exposure on the outcome is homogeneous in the population, as well as linear in the exposure (this is automatically satisfied if the exposure is binary), then the association of the instrumental variable with the outcome will always be a linear multiple of the association of the instrumental variable with the exposure. The ratio of the two associations (the instrumental variable estimate) will be the causal effect. Hence, if the causal effect is linear and homogeneous, then the instrumental variable estimate will always have the same sign (and magnitude) as the causal effect.

3.3 Monotonicity and the counterfactual instrumental variable assumptions

Under the monotonicity assumption and with a binary exposure, the instrumental variable estimate has been demonstrated to be an average of the individual causal effects in the compliers [2, 25]. With a continuous exposure, the instrumental variable estimate is a weighted average of complier-averaged causal effects for different values of the exposure [27]. This means that, under the monotonicity assumption and assuming that the causal effect of the exposure on the outcome is monotone, the instrumental variable estimate will always have the same sign as the causal effect (the average effect in the compliers). However, the cited papers used the counterfactual assumptions to define an instrumental variable. We proceed to demonstrate by counterexample that either if the monotonicity assumption is violated (Sections 3.4 and 3.5) or else if the instrumental variable only satisfies the graphical assumptions and not the counterfactual assumptions (Section 3.6), then the instrumental variable estimate and the causal effect may have different signs. Plausibility of the monotonicity assumption is discussed in Section 4.2.

3.4 Effect heterogeneity and non-monotonicity

Effect heterogeneity means that the individual causal effects differ between individuals in the population. A simple mechanism by which the direction of the instrumental variable–exposure association may not reflect the direction of the causal effect is non-monotonicity combined with effect heterogeneity.

If the population consists of 60 % compliers and 40 % defiers, and if the causal effect is 5 units in the compliers and 10 units in the defiers, then the estimated association of the exposure with the instrumental variable (in a large sample, hence the hats to denote estimates are dropped) will be $\mathbb{E}[X|(Z=1) - X|(Z=0)] = 0.6 - 0.4 = 0.2$, and the estimated association of the outcome with the instrumental variable will be $\mathbb{E}[Y|(Z=1) - Y|(Z=0)] = 0.6 \times 5 + 0.4 \times -10 = 3 - 4 = -1$. These calculations are illustrated further in Table 1. Hence both the association between the instrumental variable and the outcome and the instrumental variable estimate are negative, but the causal effect for all individuals is positive [2].

Table 1: Example 1: effect heterogeneity and non-monotonicity. Instrumental variable estimate is $\frac{3-4}{0.6-0.4} = -5$, despite positive causal effect for all individuals.

	Stratum	Expected value of exposure	Expected value of outcome
$Z=1$	Compliers	1	5
	Defiers	0	0
	Overall	$0.6 \times 1 + 0.4 \times 0 = 0.6$	$0.6 \times 5 + 0.4 \times 0 = 3$
$Z=0$	Compliers	0	0
	Defiers	1	10
	Overall	$0.6 \times 0 + 0.4 \times 1 = 4$	$0.6 \times 0 + 0.4 \times 10 = 4$

3.5 Non-linearity and non-monotonicity

A similar phenomenon can be induced if the exposure–outcome relationship is non-linear. For instance, we consider $Y(x) = \log x$, and $X|(Z=z) = 10 + \alpha z$ for $z = 0, 1$, with $\alpha = 1$ for 90 % of the population and $\alpha = -8$ for the remaining 10 % of the population. It is noted that there is no confounding or effect modification in the exposure–outcome relationship, and the causal effect of the exposure on the outcome is positive for all values of the exposure ($X > 0$). The association of the instrumental variable with the exposure is positive:

Table 2: Example 2: non-linearity and non-monotonicity. Instrumental variable estimate is $\frac{2.227-2.302}{10.1-10} < 0$, despite positive causal effect for all individuals (the outcome is an increasing function of the exposure).

	Stratum	Expected value of exposure	Expected value of outcome
$Z=1$	$\alpha=1$	11	$\log 11$
	$\alpha=-8$	2	$\log 2$
	Overall	$0.9\times 11+0.1\times 2=10.1$	$0.9\times \log 11+0.1\times \log 2=2.227$
$Z=0$	$\alpha=1$	10	$\log 10$
	$\alpha=-8$	10	$\log 10$
	Overall	$0.9\times 10+0.1\times 10=10$	$0.9\times \log 10+0.1\times \log 10=2.302$

$\mathbb{E}[X|(Z=1)-X|(Z=0)]=0.9\times 1+0.1\times -8=0.1$. However, the association of the instrumental variable with the outcome is negative: $\mathbb{E}[Y|(Z=1)-Y|(Z=0)]=(0.9\log 11+0.1\log 2)-(1\log 10)=2.227-2.302=-0.075<0$. These calculations are illustrated further in Table 2.

3.6 Non-linearity and monotonicity: Simpson's paradox

Finally, we give an example in which the direction of the instrumental variable–outcome association, causal effect, and instrumental variable estimate have different signs despite the monotonicity assumption being satisfied. We assume that the population divides into two groups $M=0$ and $M=1$, such that $\mathbb{P}(Z=0,M=0)=0.1$, $\mathbb{P}(Z=0,M=1)=0.4$, $\mathbb{P}(Z=1,M=0)=0.4$, and $\mathbb{P}(Z=1,M=1)=0.1$. We assume $X|(M=m,Z=z)=2M+Z+1$, and $Y(x)=f(x)$, where initially $f(x)=\exp x$.

Note that although Z satisfies the graphical criteria for an instrumental variable (as given in Ref. [15]), it fails the counterfactual criteria for an instrumental variable (as given in Ref. [28], Theorem 4.4.1) as it is associated with the counterfactuals $X(z)$ for $z=0,1$.

Now we have that the association of the instrumental variable with the outcome is positive: $\mathbb{E}[Y|(Z=1)-Y|(Z=0)]=(0.2\times \exp 4+0.8\times \exp 2)-(0.8\times \exp 3+0.2\times \exp 1)=16.83-16.61=0.22$, but the association of the instrumental variable with the exposure is negative: $\mathbb{E}[X|(Z=1)-X|(Z=0)]=(0.2\times 4+0.8\times 2)-(0.8\times 3+0.2\times 1)=2.4-2.6=-0.2$. Hence the instrumental variable estimate is negative. This is an example of Simpson's paradox: the direction of association between two variables does not necessarily reflect the direction of the causal effect (even in the absence of confounding, as is the case here) [29]. These calculations are illustrated further in Table 3.

Table 3: Example 3: non-linearity and monotonicity. Instrumental variable estimate is $\frac{16.83-16.61}{2.4-2.6} < 0$, despite positive causal effect for all individuals (the outcome is an increasing function of the exposure).

	Stratum	Expected value of exposure	Expected value of outcome
$Z=1$	$M=1$	4	$\exp 4$
	$M=0$	2	$\exp 2$
	Overall	$0.2\times 4+0.8\times 2=2.4$	$0.2\times \exp 4+0.8\times \exp 2=16.83$
$Z=0$	$M=1$	3	$\exp 3$
	$M=0$	1	$\exp 1$
	Overall	$0.8\times 4+0.2\times 2=2.6$	$0.8\times \exp 3+0.2\times \exp 1=16.61$

An example where there is Simpson's paradox in the instrumental variable–outcome association but not the instrumental variable–exposure association is harder to find; it requires the derivative of the exposure–outcome function to be concave. If we take $Y(x)=x\log x-x$, an increasing function in x for $x\geq 1$, and change the probabilities to $\mathbb{P}(Z=0,M=0)=0.13$, $\mathbb{P}(Z=0,M=1)=0.37$, $\mathbb{P}(Z=1,M=0)=0.37$, and

Table 4: Example 4: non-linearity and monotonicity. Instrumental variable estimate is $\frac{-0.052-(-0.041)}{2.52-2.48} = \frac{-0.011}{0.04} < 0$, despite positive causal effect for all individuals (the outcome is an increasing function of the exposure).

	Stratum	Expected value of exposure	Expected value of outcome
$Z=1$	$M=1$	4	$4\log 4-4$
	$M=0$	2	$2\log 2-2$
	Overall	$0.26\times 4+0.74\times 2=2.52$	$0.26\times(4\log 4-4)+0.74\times(2\log 2-2)=-0.052$
$Z=0$	$M=1$	3	$3\log 1-3$
	$M=0$	1	$1\log 1-1$
	Overall	$0.74\times 3+0.26\times 1=2.48$	$0.74\times(3\log 3-3)+0.26\times(1\log 1-1)=-0.041$

$\mathbb{P}(Z=1, M=1)=0.13$, then the association of the instrumental variable with the exposure is positive: $\mathbb{E}[X|(Z=1)-X|(Z=0)]=(0.26\times 4+0.74\times 2)-(0.74\times 3+0.26\times 1)=2.52-2.48=0.04$; but the association of the instrumental variable with the outcome is negative: $\mathbb{E}[Y|(Z=1)-Y|(Z=0)]=-0.052-(-0.041)=-0.011$. These calculations are illustrated further in Table 4.

Box 1: Summary of results

Sufficient conditions under which the instrumental variable and causal estimates have the same sign:
- Monotonicity of the instrumental variable–exposure association and the counterfactual instrumental variable assumptions
- Linearity and homogeneity of the exposure–outcome relationship and the graphical instrumental variable assumptions
- Homogeneity of the exposure–outcome relationship and the counterfactual instrumental variable assumptions

Scenarios under which the instrumental variable and causal estimates may have different signs:
- Effect heterogeneity and non-monotonicity
- Non-linearity and non-monotonicity
- Non-linearity and monotonicity (under the graphical instrumental assumptions)

4 Discussion

The findings of this paper are summarized in Box 1. For scenarios where the causal effect is monotone in the population, we have demonstrated that the instrumental variable estimate has the same sign as the causal effect if the causal effect is homogeneous and linear, or if the instrumental variable satisfies both the counterfactual instrumental variable assumptions and the monotonicity assumption. The first condition (homogeneity and linearity) is sufficient to estimate an average causal effect, while the second condition (monotonicity) is sufficient to estimate a local average causal effect. If these assumptions are weakened, then the instrumental variable estimate may differ in sign from the causal effect. Hence, if a researcher is willing to make the instrumental variable assumptions, but is unwilling to make additional assumptions necessary for estimating a causal effect (such as monotonicity of the instrumental variable–exposure association, or homogeneity and linearity of the exposure–outcome relationship), then they are able to conclude that the exposure has a causal effect on the outcome, but they are almost never (see "bounds" below) able to make any definite conclusion about the direction of such an effect.

4.1 Bounds for the causal effect

When the exposure is binary, the core instrumental variable assumptions only identify bounds for the causal effect rather than a single causal estimate [30, 31]. It may in some cases that these bounds are able to

determine the direction of the causal effect. However, these bounds are often extremely imprecise, and even in large samples are rarely informative. With a continuous exposure, even bounds for the causal effect cannot be made without further assumptions.

4.2 Plausibility of the monotonicity assumption

Although the monotonicity assumption is reasonably plausible in the context of randomized trials, it may be violated in other contexts, such as in Mendelian randomization. A genetic variant may have a different direction of association with the exposure in substrata of the population. This has been previously documented as "the flip-flop phenomenon" [32]. This may be due to a gene–environment interaction (for instance, the direction of association is different in smokers and non-smokers). Or else it may be due to varying genetic architectures in different ethnic groups (for instance, the variant may be correlated with one functional mutation in one group, but with a different functional mutation in another group).

Another situation where the monotonicity assumption may be violated is if the instrumental variable is an allele score, particularly an unweighted allele score [33]. An allele score is a simple way of summarizing multiple genetic variants into a univariable score by simply summing the number of exposure-increasing alleles. A weighted score can be obtained by multiplying the number of exposure-increasing alleles for each genetic variant by a weight, and then calculating the weighted sum. If the genetic variants are all instrumental variables, then the score will be an instrumental variable. The use of allele scores is motivated by the desire to avoid bias from weak instruments that can lead to misleading findings if the genetic variants do not explain much variation in the exposure [34]. As genetic variants will generally have different magnitudes of association with the exposure, the monotonicity assumption is likely to be violated for an unweighted score as greater values of the score will not necessarily correspond to greater expected values of the exposure. Even with a weighted score, violation of the monotonicity assumption is likely if the weights are not precisely estimated. Additionally, even if the weights are precisely estimated, but the associations of genetic variants vary between individuals, then the allele score may not satisfy the monotonicity assumption even if all the constituent variants individually satisfy the monotonicity assumption.

4.3 Relationship to other methodological areas

The Simpson's paradox phenomenon observed in this paper for instrumental variable analysis has also been observed for mediation analysis. Imai et al. demonstrated that it is possible for the association of an exposure with a mediator to be positive, and for the association of the mediation with the outcome to be positive, but for the average indirect effect of the exposure on the outcome to be negative [35].

4.4 Conclusion

Although situations in which the instrumental variable–outcome association has a different direction to the causal effect of the exposure on the outcome are likely to be uncommon, they are possible, particularly if the direction of effect of the instrumental variable on the exposure differs between individuals (non-monotonicity). However, even if the monotonicity assumption is satisfied, it is possible that the instrumental variable estimate and causal effect have different signs when the exposure–outcome relationship is non-linear. Hence, a definite conclusion that the exposure has (say) a positive direction of effect on the outcome requires additional assumptions to the standard instrumental variable assumptions, assumptions that are similar to those required to estimate a causal effect.

In the context of Mendelian randomization, this means that not only should the magnitude of a causal estimate not be overinterpreted [10, 36], but even the direction of the causal estimate may be a false guide as to whether the exposure should be increased or decreased unless further assumptions are made.

Funding: Wellcome Trust, (100114) Directorate for Social, Behavioral and Economic Sciences.

References

1. Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66: 688–701. DOI:10.1037/h0037350.
2. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. J Am Stat Assoc 1996;91:444–55. DOI:10.2307/2291629.
3. Lawlor D, Harbord R, Sterne J, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27:1133–63. DOI:10.1002/sim.3034.
4. Burgess S, Thompson SG. Mendelian randomization: methods for using genetic variants in causal estimation. Chichester, UK: Chapman & Hall, 2015.
5. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Meth Med Res 2007;16:309–30. DOI:10.1177/0962280206077743.
6. VanderWeele T, Tchetgen Tchetgen E, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. Epidemiology 2014;25:427–35. DOI:10.1097/ede.0000000000000081.
7. Sussman J, Wood R, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. Br Med J 2010;340:c2073. DOI:10.1136/bmj.c2073.
8. Swanson S, Hernán M. Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology 2013;24:370–4. DOI:10.1097/ede.0b013e31828d0590.
9. Hernán M, Robins J. Instruments for causal inference: an epidemiologist's dream? Epidemiology 2006;17:360–72. DOI:10.1097/01.ede.0000222409.00878.37.
10. Burgess S, Butterworth A, Malarstig A, Thompson S. Use of Mendelian randomisation to assess potential benefit of clinical intervention. Br Med J 2012;345:e7325. DOI:10.1136/bmj.e7325.
11. Ference BA, Yoo W, Alesh I, Mahajan N, Mirowska KK, Mewada A, et al. Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. J Am Coll Cardiol 2012;60:2631–9. DOI:10.1016/j.jacc.2012.09.017.
12. Taylor F, Ward K, Moore T, Burke M, Davey Smith G, Casas J, et al. Statins for the primary prevention of cardiovascular disease. Cochrane Database Syst Rev 2013;2013:1.
13. Burgess S, Butterworth A, Thompson S. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658–65. DOI:10.1002/gepi.21758.
14. Plenge R, Scolnick E, Altshuler D. Validating therapeutic targets through human genetics. Nat Rev Drug Discov 2013;12: 581–94. DOI:10.1038/nrd4051.
15. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29:722–9. DOI:10.1093/ije/29.4.722.
16. Martens E, Pestman W, de Boer A, Belitser S, Klungel O. Instrumental variables: application and limitations. Epidemiology 2006;17:260–7. DOI:10.1097/01.ede.0000215160.88317.cb.
17. Didelez V, Meng S, Sheehan N. Assumptions of IV methods for observational epidemiology. Stat Sci 2010;25:22–40. DOI:10.1214/09-sts316.
18. Clarke PS, Windmeijer F. Instrumental variable estimators for binary outcomes. J Am Stat Assoc 2012;107:1638–52. DOI:10.1080/01621459.2012.734171.
19. Joffe M, Small D, Brunelli S, Feldman H. Extended instrumental variables estimation for overall effects. Int J Biostat 2008;4:1–20. DOI:10.2202/1557–4679.1082.
20. Cox D. Planning of experiments. Section 2: some key assumptions. Chichester, UK: Wiley, 1958.
21. VanderWeele T, Hernán M. Causal inference under multiple versions of treatment. J Causal Inference 2013;1:1–20. DOI:10.1515/jci–2012–0002.
22. Frangakis C, Rubin D. Principal stratification in causal inference. Biometrics 2002;58:21–9. DOI:10.1111/j.0006-341X.2002. 00021.x.
23. Robins JM. Health service research methodology: a focus on AIDS, chap. The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. Washington, DC, USA: National Center for Health Services Research, 1989:113–59.
24. Wald A. The fitting of straight lines if both variables are subject to error. Ann Math Stat 1940;11:284–300.

25. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. Econometrica 1994;62:467–75. DOI:10.2307/2951620.
26. Allison DB, Faith MS, Heo M, Kotler DP. Hypothesis concerning the U-shaped relation between body mass index and mortality. Am J Epidemiol 1997;146:339–49.
27. Angrist J, Graddy K, Imbens G. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. Rev Econ Stud 2000;67:499–527. DOI:10.1111/1467-937x.00141.
28. Angrist J, Pischke J. Mostly harmless econometrics: an empiricist's companion. Chapter 4: Instrumental variables in action: sometimes you get what you need. Princeton, NJ, USA: Princeton University Press, 2009.
29. Pearl J. Causality: models, reasoning, and inference. Chapter 6: Simpson's paradox, confounding and collapsibility. Cambridge, UK: Cambridge University Press, 2000.
30. Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance. J Am Stat Assoc 1997;92:1171–6. DOI:10.1080/01621459.1997.10474074.
31. Chesher A. Nonparametric identification under discrete variation. Econometrica 2005;73:1525–50. DOI:10.1111/j.1468-0262.2005.00628.x.
32. Lin PI, Vance JM, Pericak-Vance MA, Martin ER. No gene is an island: the flip-flop phenomenon. The Am J Hum Genet 2007;80:531–8. DOI:10.1086/512133.
33. Burgess S, Thompson S. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol 2013;42:1134–44. DOI:10.1093/ije/dyt093.
34. Davies N, von Hinke Kessler Scholder S, Farbmacher H, Burgess S, Windmeijer F, Davey Smith G. The many weak instrument problem and Mendelian randomization. Stat Med 2015;34:454–68. DOI:10.1002/sim.6358.
35. Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. Am Political Sci Rev 2011;105:765–89. DOI:10.1017/s0003055411000414.
36. Burgess S, Butterworth AS, Thompson JR. Beyond Mendelian randomization: how to interpret evidence of shared genetic predictors. J Clin Epidemiol 2015. DOI:10.1016/j.jclinepi.2015.08.001.