Abstract

Background: Instrumental variables (IVs) can be used to provide evidence as to whether a treatment X has a causal effect on an outcome Y. Even if the instrument Z satisfies the three core IV assumptions of relevance, independence and the exclusion restriction, further assumptions are required to identify the average causal effect (ACE) of X on Y. Sufficient assumptions for this include: homogeneity in the causal effect of X on Y; homogeneity in the association of Z with X; and no effect modification (NEM).

Methods: We describe the NO Simultaneous Heterogeneity (NOSH) assumption, which requires the heterogeneity in the X-Y causal effect to be mean independent of (i.e., uncorrelated with) both Z and heterogeneity in the Z-X association. This happens, for example, if there are no common modifiers of the X-Y effect and the Z-X association, and the X-Y effect is additive linear. We illustrate NOSH using simulations and by re-examining selected published studies.

Results: When NOSH holds, the Wald estimand equals the ACE even if both homogeneity assumptions and NEM (which we demonstrate to be special cases of – and therefore stronger than – NOSH) are violated.

Conclusions: NOSH is sufficient for identifying the ACE using IVs. Since NOSH is weaker than existing assumptions for ACE identification, doing so may be more plausible than previously anticipated.

Keywords: Causal inference; Effect modification; Homogeneity; Identification; Instrumental variables.

1. Introduction

Instrumental variable (IV) methods can be applied to strengthen causal inference using non-experimental data when the assumption of no unmeasured confounding is implausible.[1–3] A valid IV satisfies the three core assumptions (defined formally below) of relevance, independence and the exclusion restriction.[4,5] These assumptions allow testing the sharp causal null hypothesis – i.e., whether the treatment has a causal effect on the outcome for at least one member of the population. However, point-estimating well-defined treatment effect parameters requires further assumptions, sometimes referred to as fourth point identifying assumptions (IV4).[6] In this paper our target estimand is the average causal effect (ACE) on the additive scale (defined precisely in section 2.1). Different IV4 assumptions sufficient for ACE identification have been proposed, including: treatment effect homogeneity,[4,7] instrument effect homogeneity,[8,9] no effect modification (NEM)[4,9], and no unmeasured common effect modifier (NUCEM)[10] (defined precisely in section 2.3). Given our focus on the ACE, IV4 assumptions that identify other estimands (e.g., monotononicity, which identifies the local average causal effect (LACE)[11]) will not be discussed in detail.

Previous papers have described the IV estimand when treatment effects vary. Heckman (1997) described, for a binary treatment, that if individual treatment effects do not affect whether someone is treated, the IV estimand equals the ACE.[7] Harris and Remler (1998) noted this, stating that the ACE is identified if treatment effect heterogeneity is unrelated to treatment assignment.[12] Brookhart and Schneeweiss (2007) described how, if the individual levels effects of a binary treatment are the same for different instrument values, the IV estimator identifies the ACE in the population.[8] Wang and Tchetgen Tchetgen (2018) describe how, for a binary instrument and a binary treatment, the ACE is identified if no unmeasured confounders are additive modifiers of the association between the instrument and the treatment or of the effect of the treatment on the outcome.[9] Cui and Tchetgen Tchetgen (2021), again focusing on binary instrument and treatment, proposed a weaker version of this condition which only requires that there is no common additive modifier of the instrument-treatment association and the treatment-outcome effect.[10] Syrgkanis et al. (2019) state that the ACE

is not generally identified if individual-level instrument-treatment and treatment-outcome effects are dependent.[13]

Numerous reviews and methodological papers have described several identifying assumptions.[4,14–17] Here, we introduce the NO Simultaneous Heterogeneity (NOSH) assumption. We show that if NOSH holds, the IV estimand equals the ACE. We also show that other IV4 assumptions are special cases of NOSH. Finally, we use simulations to corroborate the theory.

2. Methods

2.1. Notation and assumptions

Let Z, X, Y and U respectively denote the instrument, the time-fixed treatment, the outcome and all unmeasured common causes of X and Y. For simplicity, we consider the case where no adjustment is made for measured covariates. However, the concepts developed here can be trivially extended to accommodate measured covariates. We also discuss covariate adjustment in the simulation study (section 4 in the Supplement).

Z is a valid IV if the following three causal assumptions (illustrated in Figure 1) are satisfied: i) relevance: Z and X are statistically dependent; ii) independence: $Z \perp\!\!\!\perp U$ (where "$\perp\!\!\!\perp$" denotes statistical independence); and iii) exclusion restriction: $Z \perp\!\!\!\perp Y|X, U$.[4,5]

Figure 1 is a graphical representation of the following non-parametric structural equation model:

$$X_i = f_X\big(Z = Z_i, U = U_i, \varepsilon_X = \varepsilon_{X_i}\big)$$

$$Y_i = f_Y\big(X = X_i, U = U_i, \varepsilon_Y = \varepsilon_{Y_i}\big),$$

where X_i is the value of X for individual i (the same notation applies to other variables), f_X and f_Y respectively denote the functions governing X and Y, and ε_X and ε_Y respectively denote stochastic direct causes of X and Y (so that $\varepsilon_X \perp\!\!\!\perp \varepsilon_Y$). We now define $F_{X_i}(z) = \mathrm{E}\big[X_i|do(Z = z), U = U_i, \varepsilon_X = \varepsilon_{X_i}\big]$ and $F_{Y_i}(x) = \mathrm{E}\big[Y_i|do(X = x), U = U_i, \varepsilon_X = \varepsilon_{X_i}\big]$ – that is, $F_{X_i}(z)$ is the expectation of X when Z

is set (possibly counterfactually) to z, while the other variables retain their observed values. A similar interpretation holds for $F_{Y_i}(x)$.

The individual level instrumental and treatment effects β_X and β_Y can be defined as follows:

$\beta_{X_i} = F_{X_i}(1) - F_{X_i}(0)$ for a binary Z,

$\beta_{X_i} = F_{X_i}(z) - F_{X_i}(z-1)$ for a multivalued discrete Z,

$\beta_{X_i} = \frac{\partial}{\partial z}\left[F_{X_i}(z)\right]\Big|_{z=Z_i}$ for a continuous Z,

$\beta_{Y_i} = \mathrm{E}\left[F_{Y_i}(1) - F_{Y_i}(0) | U = U_i, \varepsilon_Y = \varepsilon_{Y_i}\right]$ for a binary X,

$\beta_{Y_i} = \frac{\partial}{\partial x}\left[F_{Y_i}(x)\right]\Big|_{x=X_i}$ for a continuous X.

For a multivalued discrete Z, we assume Z is coded numerically such that $\mathrm{E}[X|Z=1] \leq \cdots \leq \mathrm{E}[X|Z=K]$, where K is the number of values that Z attains, and $z \in \{2, \ldots, K\}$. Notice that, for continuous Z, the definition of β_X implicitly assumes that $F_X(z)$ is differentiable with respect to Z. For non-continuous X, this would happen for example if $X_i \sim \mathrm{Bernoulli}(p_i)$ or $X_i \sim \mathrm{Poisson}(\lambda_i)$, where p_i or λ_i are differentiable functions of Z. A similar notion applies to β_Y and $F_Y(z)$ for a continuous X.

Under the stable unit treatment value assumption, β_{X_i} is the additive change in the expectation of X caused by a unit increase in Z in individual i and β_{Y_i} is the additive change in the expectation of Y caused by a unit increase in X in individual i. From this notation, the ACE is defined as $\mathrm{E}[\beta_Y]$ – i.e., the average of β_Y in the population. This definition incorporates the case of a multi-valued X (excluding the case where X is an unordered multivalued variable or X is a discrete variable with non-linear effects on Y, since derivatives are not defined in these cases), where the distribution of X in the population will affect the ACE. This quantity is sometimes referred to as the average derivative effect.[18,19] Finally, the conventional IV estimand known as the Wald estimand (here denoted as β_{IV}) is defined as $\beta_{IV} = \frac{\mathrm{cov}(Y,Z)}{\mathrm{cov}(X,Z)}$.[13,20]

2.2. The NO Simultaneous Heterogeneity (NOSH) assumption

We define the NOSH assumption as a combination of two conditions (Assumptions 1 and 2, defined below). The name NOSH refers to the fact that, if these two assumptions hold, then $\beta_Y \perp\!\!\!\perp (Z, \beta_X)$ – that is, heterogeneity in the causal effect is independent of heterogeneity in the instrument effect (and of the instrument). Since β_X and β_Y denote effects on the additive scale, NOSH focuses on additive effect modification. Of note, there is no assumption regarding multiplicative effect modification other than what is implied by our assumptions on additive effect modification (in section 2 in the supplement, we discuss implications of non-linear effects and non-linear data-generating models for NOSH).

Theorem 1: If NOSH holds, then $\beta_{IV} = \mathrm{ACE}$ (proof in section 3 in the Supplement).

We now define the conditions for NOSH to hold using causal diagrams (see section 1 in the Supplement for equivalent definitions using non-parametric structural equation models). For a precise articulation, it is useful to partition U in Figure 1 in six non-overlapping, exhaustive sets of variables (Table 1).

Figure 2A illustrates possible causal relationships compatible with Figure 1 (i.e., compatible with the core IV assumptions) among U, Z, X, β_X and β_Y. Of note, Figure 1 assumes that Z is a causal instrument, but this is not necessary for NOSH to be defined or hold. In Theorem 1, β_X can be replaced β_X^*, here denoting the individual-level association of a non-causal instrument Z^* and X. By non-causal instrument, we mean that Z^* is not a cause of X. Still, it is associated with X through paths that included Z as a non-collider (more specifically, paths of the form $Z^* \leftarrow W \rightarrow Z$ or $Z^* \rightarrow \boxed{C} \leftarrow Z$, where C is being conditioned on). Therefore, both modifiers of the effect of Z on X and modifiers of the association between Z^* and Z will be modifiers of the association of Z^* and X. Modifiers of Z-X must be independent of modifiers of X-Y for NOSH to hold, as discussed above. Moreover, modifiers of Z-Z^* must be independent of any cause of Y (otherwise Z^* would be an

invalid IV). Therefore, if Assumptions 1 and 2 (see below) hold for Z, they also hold for Z^*. Given this clarification, Z will be depicted as a causal instrument in Figure 2 for simplicity.

Although it is not usual to depict individual-level effects as nodes in a causal graph, β_X and β_Y are indeed random variables. Since X is fixed in time, these two variables are simply functions of other individual-level random variables, so they are not qualitatively different from X or Y for example. This contrasts with variables that may also depend on non-individual level characteristics, such as person-time, which depends on follow-up duration and can be influenced by study design. Causes of β_X can be interpreted as modifiers of the effect of Z on X, while causes of β_Y can be interpreted as modifiers of the effect of X on Y. A more comprehensive description on representing individual-level effects in causal graphs is available elsewhere.[21] For example, Table 1 describes that U_2 modifies the effect of X on Y, but not the effect of Z on X. This is translated in Figure 2A by the directed path from U_2 to β_Y, but not to β_X (see Box 1). Explicitly depicting β_X and β_Y in the graph facilitates identifying conditions under which β_X and β_Y are d-separated, which is useful to precisely articulate sufficient conditions for NOSH to hold.

In Figure 2A, all unmeasured variables U_1 to U_6 are d-connected to one another due to a latent variable, thus allowing for statistical dependencies between them in unrestricted ways. Although these variables could be d-connected due to other causal structures (e.g., a path $U_1 \rightarrow U_2$), this would violate the classification proposed in Table 1 (in this example, U_1 would be an effect modifier of both the effect of Z on X and the effect of X on Y – i.e., it would be a component of U_6), and thus blur the distinct implications of distinct types of effect modifiers.

Although this is not central to our arguments, it is instructive to clarify that, from a statistical perspective, both U_3 and U_4 (for example) are effect modifiers of β_X if U_3 and U_4 are d-connected, because β_X will vary between strata of U_3 and between strata of U_4 (the same reasoning applies to other unmeasured variables with respect to β_X and/or β_Y). For clarity, we use "effect modifier" to refer to variables that themselves exert effect modification and "surrogate effect modifier" to refer

to variables that are d-connected with effect modifiers but do not themselves exert effect modification. Although our arguments could ignore surrogate effect modifiers, defining different types of effect modifiers allows for a more comprehensive articulation of conditions sufficient for NOSH to hold.

Since d-separation implies statistical independence, causal diagrams can be used to find conditions under which β_Y is d-separated from (and thus statistically independent of) β_X and Z. Under such conditions, NOSH holds. In Figure 2A, β_Y is d-connected through multiple paths to both Z and β_X, so $\beta_Y \perp\!\!\!\perp (Z, \beta_X)$ (and therefore NOSH) will not generally hold.

We now describe the assumptions that define NOSH.

Assumption 1: All unmeasured variables that modify Z-X are independent of all unmeasured variables that modify X-Y.

This assumption implies that there are no unmeasured variables that themselves modify (on the additive scale) both the effect of Z on X and the effect of X on Y – that is, $U_6 = \emptyset$. Furthermore, all unmeasured variables that modify the effect of Z on X (i.e., U_1 and U_4) are independent of all unmeasured variables that modify the effect of X on Y (i.e., U_2 and U_5). This implies that there are neither unmeasured effect modifiers of Z-X that are also surrogate effect modifiers of X-Y nor unmeasured surrogate effect modifiers of Z-X that are also effect modifiers of X-Y. However, U_1, U_3 and U_4 may be correlated with one another, and U_2, U_3 and U_5 may be correlated with one another. For this to hold, X cannot cause U_2. Otherwise, U_1 and U_4 (which are modifiers of the effect of Z on X) and U_2 (which is a modifier of the effect of X on Y) will be d-connected through the paths $U_1 \rightarrow X \rightarrow U_2$ and $U_4 \rightarrow X \rightarrow U_2$. This assumption is violated in Figure 2A, where all unmeasured variables are allowed to be d-connected with one another.

Even if Assumption 1 holds, paths of the form $\beta_X \leftarrow Z \rightarrow X \rightarrow \beta_Y$, which render β_X and β_Y d-connected, may still exist. Therefore, Assumption 1 is necessary, but not sufficient, to render β_X and β_Y d-separated.

Assumption 2: The effect of X on the expectation of Y is additive linear.

This assumption holds if $F_{Y_i}(x) = \varphi_i + \mu_i x$, where φ_i and μ_i may vary between individuals. In this case, $F_{Y_i}(x) - F_{Y_i}(x') = \mu_i(x - x')$. This implies that, for a given individual, the additive change in the expectation of Y caused by a unit increase in X does not depend on the value of X that the individual has. That is, the effect of X on Y is additive linear (but not necessarily constant) across individuals. This implies that the path $X \rightarrow \beta_Y$ does not exist. Consequently, Z and β_Y are d-separated (that is, $\beta_Y \perp\!\!\!\perp Z$). Of note, this assumption is automatically satisfied if X is binary.

In Figure 2B, both Assumptions 1 and 2 hold. In this graph, β_Y and β_X are d-separated (because all paths from β_X to β_Y contain at least one collider), and β_Y and Z are d-separated (because all paths from Z to β_Y contain X as a collider).

Even though the above focused on the structural interpretation of NOSH (i.e., an interpretation where concepts can be represented in causal graphs), this assumption can be relaxed in the sense that it does not require full independence, but only mean independence (i.e., uncorrelatedness). That is, if $\mathrm{E}[\beta_Y|Z, \beta_X] = \mathrm{E}[\beta_Y]$, then $\beta_{IV} = \mathrm{ACE}$ (see the proof in the Supplement for details). Therefore, NOSH is a statistical statement, so representing it using causal graphs may be useful, but not strictly required. For example, In Figure 2A, even though NOSH will not generally hold, it is possible that it holds under lack of faithfulness (i.e., when there is d-connection but no statistical association). In this sense, NOSH is agnostic to whether faithfulness is assumed.

2.3. NOSH is a generalization of previous IV4 assumptions

We now show that well-known assumptions that identify the ACE imply NOSH.

2.3.1. Causal effect homogeneity

For a binary X and assuming deterministic counterfactuals, this assumption can be defined as $Y_i(X_i = 1) - Y_i(X_i = 0) = c$ (a constant), where $Y_i(X_i = x) = f_Y\big(do(X = x), U = U_i, \varepsilon_Y = \varepsilon_{Y_i}\big)$ for $x \in \{0,1\}$. More generally, this condition can be defined as $\beta_{Y_i} = c$. Since β_Y is constant, NOSH

trivially holds – i.e., NOSH is implied by causal effect homogeneity. Moreover, NOSH allows identification when there is causal effect heterogeneity. Therefore, NOSH is weaker than causal effect homogeneity.

2.3.2. Instrument effect homogeneity

For a binary causal instrument Z and assuming deterministic counterfactuals, this assumption can be defined as $X_i(Z_i = 1) - X_i(Z_i = 0) = c$, where $X_i(Z_i = Z) = f_X\big(do(Z = z), U = U_i, \varepsilon_X = \varepsilon_{X_i}\big)$ for $x \in \{0,1\}$. Generally, this condition can be defined as $\beta_{X_i} = c$. Since β_X is constant, $\beta_Y \perp\!\!\!\perp \beta_X$ trivially holds. However, instrument effect homogeneity does not imply $\beta_Y \perp\!\!\!\perp Z$, because the effect of X on Y may be non-linear, except if X is binary. This is important because additive linearity in the effect of X on Y is required for the conventional IV estimand to equal the ACE[22] (see section 3 in the Supplement).

Therefore, instrument effect homogeneity implies that Assumption 1 is true. However, ACE identification also requires Assumption 2, which means NOSH holds. Since NOSH allows identification under instrument effect heterogeneity and both NOSH and instrument effect homogeneity require Assumption 2, NOSH is weaker than instrument effect homogeneity.

2.3.3. No effect modification (NEM) by unmeasured factors

Homogeneity can be relaxed by considering a condition sometimes referred to as NEM. We consider two versions of this assumption: NEM1 and NEM2. For a binary X, NEM1 is defined as $\mathrm{E}[Y_i(X_i = 1) - Y_i(X_i = 0)|U_i] = \mathrm{E}[Y_i(X_i = 1) - Y_i(X_i = 0)]$.[9] More generally, NEM1 postulates that $\mathrm{E}\big[\beta_{Y_i}|U_i\big] = \mathrm{E}\big[\beta_{Y_i}\big]$ – that is, no unmeasured X-Y confounder modifies the additive effect of X on the expectation of Y. For a binary X, NEM1 holds if U and β_Y are d-separated; otherwise, it would not hold in general. Therefore, Assumption 1 is necessary for NEM1 to hold; otherwise, there will be open paths between β_Y and (components of) U. However, it is not sufficient since Assumption 1 allows for confounders to be effect modifiers. If X is continuous, Assumption 2 is also necessary (but not sufficient), otherwise the path $X \rightarrow \beta_Y$ will exist, which would render β_Y and U d-connected, thus

violating NEM1. Since Assumptions 1 and 2 are both necessary for NEM1 and sufficient for NOSH to hold, NEM1 implies NOSH. However, since these two assumptions are insufficient for NEM1 to hold, NOSH does not imply NEM1. Therefore, NOSH is weaker than NEM1. Of note, some authors refer to violation of NEM1 as essential heterogeneity.[23]

Although not typically referred to this way, NEM also applies to the association between Z and X (we will call this condition NEM2). For a binary causal instrument Z, NEM2 is defined as $\mathrm{E}[X_i(Z_i = 1) - X_i(Z_i = 0)|U_i] = \mathrm{E}[X_i(Z_i = 1) - X_i(Z_i = 0)]$.[9] More generally, NEM2 postulates that $\mathrm{E}[\beta_{X_i}|U_i] = \mathrm{E}[\beta_{X_i}]$ – that is, no unmeasured X-Y confounder modifies the additive association between Z and X. If both Z and X are binary, NEM2 holds if U and β_X are d-separated; otherwise, it would not hold in general. Therefore, Assumption 1 is necessary for NEM2 to hold; otherwise, there will be open paths between β_Y and (components of) U. However, it is not sufficient, since Assumption 1 allows for confounders to be effect modifiers. For a continuous X, NEM2 is not sufficient to identify the ACE since even the stronger condition of instrument effect homogeneity requires Assumption 2. Since Assumptions 1 and 2 are both necessary for NEM2 to identify the ACE and sufficient for NOSH to hold, NEM2 implies NOSH. However, since these two assumptions are insufficient for NEM2 to hold, NOSH does not imply NEM2. Therefore, NOSH is weaker than NEM2.

2.3.4. No unmeasured effect modification (NUCEM)

For binary Z and X, it has been shown that the usual IV estimand equals the ACE if $\mathrm{Cov}(\mathrm{E}[X|U, Z = 1] - \mathrm{E}[X|U, Z = 0], \mathrm{E}[Y|U, do(X = 1)] - \mathrm{E}[Y|U, do(X = 1)]) = 0$. In words, this condition (which is clearly weaker than NEM), postulates that there are no unmeasured variables that modify (on the additive scale) both the Z-X association and the X-Y effect.[10] This zero-covariance condition can be expressed as $\mathrm{E}[\beta_Y|\beta_X] = \mathrm{E}[\beta_Y]$, which is equivalent to NOSH for binary Z and X (this is because, for a binary X, $\mathrm{E}[\beta_Y|Z, \beta_X] = \mathrm{E}[\beta_Y|\beta_X]$ since in this case Assumption 2 is guaranteed to hold). Therefore, NOSH generalizes NUCEM to situations involving non-binary Z and X.

3. Simulation study

We performed a simulation to demonstrate further NOSH is sufficient for ACE identification (a detailed description is provided in section 4 in the Supplement). Results are shown in Figure 3. In scenario 1, NOSH holds, and the two-stage least squares estimator (TSLS) (equivalent to the Wald estimator for a single Z, X and Y[24,25]) consistently estimates the ACE, with coverage being approximately 95% and bias converging to zero as sample size increases. A similar pattern was seen when NOSH holds, but error terms were non-normal (scenarios 4 and 5). When NOSH is violated (scenarios 2 and 3), TSLS had substantial bias and undercoverage. Supplementary Table 1 shows the results for different TSLS specifications. When NOSH is violated, but Assumption 2 holds (as in scenario 2), adjusting for measured common effect modifiers mitigates bias. However, when NOSH is violated exclusively due to Assumption 2 being invalid (as in scenario 3), covariate adjustment does not improve estimation.

4. Discussion

This paper shows that the Wald estimator is consistent for the ACE if, in addition to the core IV assumptions, the NOSH assumption holds. NOSH is weaker than previously proposed IV4 assumptions that identify the ACE. This does not include the monotonicity assumption (defined precisely in section 5.1 in the Supplement), which is sometimes classified as an IV4 assumption but does not identify the ACE. While NOSH is not strictly weaker or stronger than monotonicity, NOSH has two advantages over monotonicity. First, the latter only identifies LACEs, whereas NOSH identifies the ACE. Second, the NOSH assumption identifies the ACE even if there are defiers. However, if NOSH is violated but monotonicity holds, IV estimators, identify the LACE, which is a well-defined causal parameter for binary treatments. However, the notion of compliers is not well-defined for continuous treatments. In this case, monotonicity allows interpreting the Wald estimand as a weighted average of treatment effects, with subgroups of the population where the Z-X association is stronger receiving greater weight.[20,26] Although mathematically well-defined, this parameter is difficult to interpret for policy making. Conversely, if NOSH holds, then IV estimators will identify the ACE of a continuous treatment.

In many recent papers describing methodological developments that relax the exclusion restriction assumption when there are multiple IVs, treatment effect homogeneity is implicitly or explicitly required.[27,28] This is because these methods assume that valid IVs identify the same causal effect (generally, the ACE). However, assuming NOSH holds for all IVs is sufficient to identify the ACE. This implies that the assumptions required for the validity of these methods are weaker than previously considered. Nevertheless, since we defined NOSH for a time-fixed treatment, caution must be taken to extrapolate our conclusions to time-varying treatments. Moreover, even though NOSH is weaker than homogeneity or NEM, it is still quite strong and should not be taken for granted in practice.

NOSH is an untestable assumption that cannot be guaranteed by study design. Therefore, assessing its plausibility requires subject matter knowledge. We illustrate this by discussing three published IV studies (section 5 in the Supplement). The possibility of (partially) empirically verifying some IV4 assumptions has implications for NOSH plausibility. For example, in the case of a continuous treatment, instrument effect heterogeneity would often imply (except in some specific circumstances) that the treatment is heteroscedastic with respect to the instrument (i.e., the variance of the treatment would differ between levels of the instrument).[29,30] Future methodological studies are required to assess the power and utility of such tests in typical IV settings. Since NOSH is weaker than homogeneity, homoscedasticity, in this case, would support (but not guarantee) that NOSH holds. However, heteroscedasticity would not necessarily imply that NOSH is violated. Brookhart (2007)[8] proposed empirically assessing the plausibility of instrument effect homogeneity by testing if instrument strength varies between strata of measured covariates. This can also be viewed as a test of NOSH since if multiple covariates modify instrument strength, then the assumption that there are no unmeasured common effect modifiers is less plausible. Indeed, assuming Z is a valid instrument, variability in IV estimates between strata defined by such covariates could be interpreted as evidence against NOSH. Strategies to empirically verify this assumption remain to be formally investigated.

An earlier glimpse of NOSH in the literature can be found, for example, in Angrist (1990),[31] who noted that variation in response to the draft (i.e., Z-X heterogeneity) based on potential outcomes (i.e., X-Y heterogeneity) could mean his results were biased estimates of the ACE. Indeed, this would be a scenario where NOSH is violated. For binary Z and X, NOSH is equivalent to NUCEM.[10] However, no causal structures (e.g., in the form of causal graphs) that dictate whether NOSH is violated were presented. Syrgkanis et al. (2019)[13] also postulated an independence condition that is equivalent to NOSH, without necessarily restricting to binary Z and X. However, mechanisms that could render this condition satisfied were similarly not discussed, and no explicit consideration was given to potential implications of non-linear instrument-outcome associations or treatment-outcome effects. Moreover, their assumed data-generating mechanism for the outcome was $Y_i = \theta_i(U = U_i)X_i + \rho(U = U_i) + e_i$, where $\theta(U)$ is the causal effect function, which was assumed to be linear additive. However, such a model may not be appropriate when Y is binary, which is often assumed to be a non-linear function of X and U (e.g., the expectation of Y may be a logistic function of X and U).

Here we use counterfactual notation to explicitly define individual-level effects in a framework that allows for both binary and continuous Z, X and/or Y. We also comprehensively describe data-generating mechanisms influencing NOSH using causal graphs. This helps to apply expert knowledge to assess the plausibility of this assumption in practice. We also explicitly consider the implications of non-linear Z-X associations and X-Y effects. The present work thus clarifies the assumptions underlying NOSH, which allows differentiating it from previous IV4 assumptions and explicitly propose NOSH as an IV4 assumption that is weaker than previously described IV4 assumptions that identify the ACE.

References

1. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiments. Journal of Economic Perspectives. 2001;15:69–85.

2. Bollen KA. Instrumental variables in sociology and the social sciences. Annu Rev Sociol. 2012;38:37–72.

3. Ertefaie A, Small DS, Flory JH, Hennessy S. A tutorial on the use of instrumental variables in pharmacoepidemiology. Pharmacoepidemiol Drug Saf. 2017;26:357–67.

4. Hernán MA, Robins JM. Instruments for causal inference: An epidemiologist's dream? Epidemiology. 2006;17:360–72.

5. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007;16:309–30.

6. Hernán MA, Robins JM. Causal Inference: What If. Boca Raton (FL, USA): Chapman & Hall/CRC; 2020.

7. Heckman J. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. Journal of Human Resources. 1997;32:441–62.

8. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: Assessing validity and interpreting results. International Journal of Biostatistics. 2007;3:Article 14.

9. Wang L, Tchetgen Tchetgen E. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. J R Stat Soc Series B Stat Methodol. 2018;80(3):531–50.

10. Cui Y, Tchetgen Tchetgen E. A Semiparametric Instrumental Variable Approach to Optimal Treatment Regimes Under Endogeneity. J Amer Statist Assoc. 2020;116(533):162–73.

11. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. J Am Stat Assoc. 1996;91:444–55.

12. Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. Health Serv Res. 1998;33:1337–1360.

13. Syrgkanis V, Lei V, Oprescu M, Hei M, Battocchi K, Lewis G. Machine Learning Estimation of Heterogeneous Treatment Effects with Instruments. NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019;15193–202. Available from: https://dl.acm.org/doi/pdf/10.5555/3454287.3455648.

14. Imbens GW. Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). J Econ Lit. 2010;48(2):399–423.

15. Aronow PM, Carnegie A. Beyond LATE: Estimation of the average treatment effect with an instrumental variable. Political Analysis. 2013;21(4):492–506.

16. Angrist JD, Fernández-Val I. ExtrapoLATE-ing: External validity and overidentification in the LATE framework. In: Acemoglu D, Arellano M, Dekel E, editors. Advances in Economics and Econometrics: Tenth World Congress. Cambridge (England): Cambridge University Press; 2011. p. 401–34.

17. Vansteelandt S, Didelez V. Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators. Scandinavian Journal of Statistics. 2018;45(4):941–61.

18. Hardle W, Stoker TM. Investigating Smooth Multiple Regression by the Method of Average Derivatives. J Am Stat Assoc. 1989;84(408):986.

19. Powell JL, Stock JH, Stoker TM. Semiparametric Estimation of Index Coefficients. Econometrica. 1989;57(6):1430.

20. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. J Am Stat Assoc. 1995;90:431–42.

21. Nilsson A, Bonander C, Strömberg U, Björk J. A directed acyclic graph for interactions. Int J Epidemiol. 2021;50(2):613–9.

22. Hartwig FP, Wang L, Smith GD, Davies NM. Homogeneity in the instrument-treatment association is not sufficient for the Wald estimand to equal the average causal effect when the exposure is continuous. Epidemiology. 2022;33(6):828–31.

23. Basu A, Heckman JJ, Navarro-Lozano S, Urzua S. Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. Health Econ. 2007;16(11):1133–57.

24. Angrist JD, Graddy K, Imbens GW. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. Review of Economic Studies. 2000;67(3):499–527.

25. Angrist JD, Pischke J-S. Chapter 4. Instrumental Variables in Action: Sometimes You Get What You Need. In: Mostly Harmless Econometrics. Princeton University Press; 2009. p. 113–220.

26. von Hinke S, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as instrumental variables. J Health Econ. 2016;45:131–48.

27. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40:304–14.

28. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46:1985–98.

29. Mills HL, Higgins JPT, Morris RW, Kessler D, Heron J, Wiles N, et al. Detecting Heterogeneity of Intervention Effects Using Analysis and Meta-analysis of Differences in Variance Between Trial Arms. Epidemiology. 2021;32(6):846–54.

30. Sanderson E, Glymour MM, Holmes M v., Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. Nature Reviews Methods Primers. 2022;2(1):1–21.

31. Angrist JD. Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. American Economic Review. 1990;80:313–36.

Table 1. Subsets of unmeasured variables collectively represented as U in Figure 1.

Variable	Causes X	Causes Y	Modifies* Z- X	Modifies* X- Y
U_1†	Yes	No	Yes	No
U_2†	No	Yes	No	Yes
U_3	Yes	Yes	No	No
U_4	Yes	Yes	Yes	No
U_5	Yes	Yes	No	Yes
U_6	Yes	Yes	Yes	Yes

†Strictly speaking, U_1 and U_2 do not belong to U in Figure 1, but we will use the notation U_j to refer to unmeasured variables in general.

*This refers to additive effect modification.

Figure 1. Causal graph illustrating the instrumental variable assumptions.

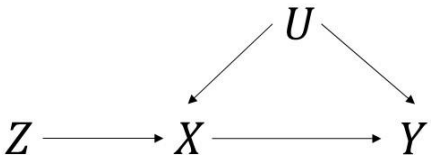
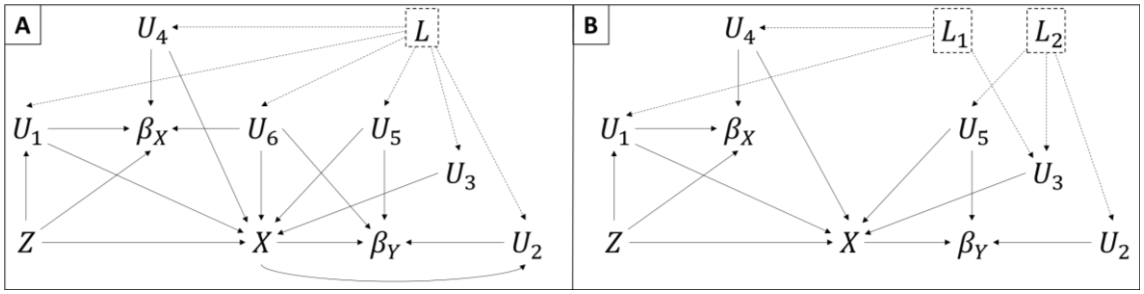


Figure 2. Causal graph illustrating causal relationships between the instrument (Z), the treatment (X), individual level causal effect of Z on X (β_X)*, individual level causal effect of X on the Y (β_Y)* and six unmeasured variables (U_1 to U_6). The dotted lines represent causal effects from unknown causes (L, L_1 and L_2, with a dotted box marking them as latent) that render some or all unmeasured variables d-connected.

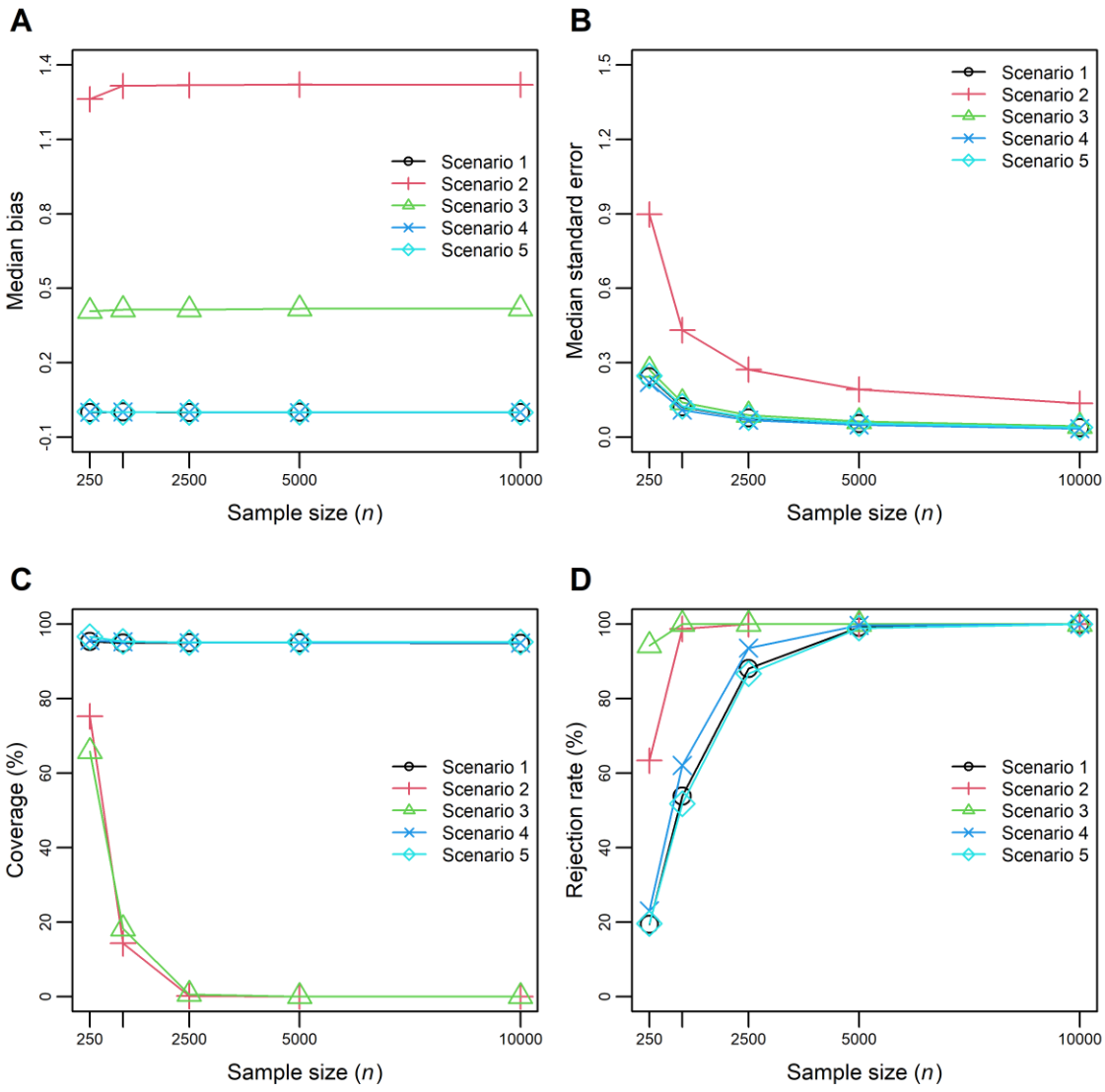


A: All possible causal relationships compatible with Figure 1 (i.e., unrestricted model, where assumptions 1 and 2 are violated).

B: Assumptions 1 and 2 hold.

*Since β_X and β_Y represent effects on the additive scale, variables d-connected with them are additive effect modifiers of Z-X and X-Y, respectively.

Figure 3. Median bias, median standard error, coverage and rejection rate of two-stage least squares as an estimator of the ACE† in scenarios 1-5‡, where the causal effect of X on Y is linear.



†Average causal effect (ACE).

‡1: NOSH holds. 2: Assumption 1 violated. 3: Assumption 2 violated. 4: NOSH holds and error terms are drawn from a beta distribution. 5: NOSH holds and error terms are drawn from a mixed chi-squared distribution.

Box 1. Illustration of the equivalence between effect modification and causes of individual level effects

Graphically representing effect modification as causes of individual level effects is consistent with the definition of effect modification. To illustrate, suppose $Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i V_i + \varepsilon_i$, where V_i is not caused by X and ε_i is some continuous random error. In this situation, β_Y is defined as follows:

$$\beta_{Y_i} = \frac{\partial}{\partial X}[\beta_{Y_i}] = \alpha_2 + \alpha_3 V_i.$$

In this example, V is a cause of β_Y because it is in the right-hand side of the structural equation model for β_Y. This would be represented in a causal graph as any other causal relationship – i.e., as a directed path from the cause (V) to the consequence (β_Y).