

1
2
3
4 **To appear in Volume 5, Issue 3 (September 2018)**
5 **SpringerLink Header: Epidemiologic Methods (R Maclehose, Section Editor)**
6 **Topical Collection on Epidemiologic Methods**
7
8
9

10 **Understanding the assumptions underlying instrumental variable analyses: A brief review**
11 **of falsification strategies and related tools**
12
13
14

15 Jeremy Labrecque¹, Sonja A. Swanson^{1,2}
16
17

18 1. Department of Epidemiology, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, NL. 2.
19 Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, US.
20

21 Corresponding author:
22 Jeremy Labrecque
23 Department of Epidemiology
24 Erasmus MC
25 P.O. Box 2040
26 3000 CA Rotterdam
27 The Netherlands
28 j.labrecque@erasmusmc.nl
29
30
31
32
33
34
35

36 Keywords: instrumental variable; falsification; Mendelian randomization
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Abstract

Purpose of review: Instrumental variable (IV) methods continue to be applied to questions ranging from genetic to social epidemiology. In the epidemiologic literature, discussion of whether the assumptions underlying IV analyses hold is often limited to only certain assumptions and even then, arguments are mostly made using subject matter knowledge. To complement subject matter knowledge, there exist a variety of falsification strategies and other tools for weighing the plausibility of the assumptions underlying IV analyses.

Recent findings: There are many tools that can refute the IV assumptions or help estimate the magnitude or direction of possible bias if the conditions do not hold perfectly. Many of these tools, including both recently developed strategies and strategies described decades ago, are underused or only used in specific applications of IV methods in epidemiology.

Summary: Although estimating causal effects with IV analyses relies on unverifiable assumptions, the assumptions can sometimes be refuted. We suggest that the epidemiologists using IV analyses employ all the falsification strategies that apply to their research question in order to avoid settings that demonstrably violate a core condition for valid inference.

Introduction

Many epidemiologists rely on, but are simultaneously skeptical of, the exchangeability (or no uncontrolled confounding) condition required to identify causal effects in our typical analyses of observational studies. Exchangeability is difficult to achieve and impossible to verify, which has led some epidemiologists to prefer instrumental variable (IV) methods that trade in this exchangeability condition for other conditions that are perceived as more plausible in some settings.

IV analysis requires first and foremost an instrumental variable. That is, it requires a variable that meets three conditions: (1) it is associated with the exposure (“relevance”), (2) it only affects the outcome through the exposure (“exclusion restriction”) and (3) its effect on the outcome is unconfounded (“exchangeability”) [1].* These are the three requisite IV conditions, although as we describe below additional conditions are necessary to identify causal effects. We can see then that, similar to the exchangeability condition in traditional epidemiologic approaches, IV analysis relies on unverifiable conditions. As IV analyses have grown in popularity in recent decades, so have concerns over the plausibility of these assumptions. The most common method of arguing that the IV assumptions hold is using substantive knowledge. However, there are many methods and tools that use the data in hand that can further strengthen or refute the IV assumptions or help estimate the magnitude or direction of possible bias if the conditions do not hold perfectly. Many of these methods are underused or only used in specific applications of IV methods in epidemiology [2].

Here, we aim to describe and discuss the tools that are available to epidemiologists to strengthen IV analyses (Table 1). Before continuing, we note that IV methods have generally been applied by epidemiologists in four settings: Mendelian randomization studies that propose genetic variants as instruments [3]; pharmacoepidemiologic studies that propose geographic, provider, or temporal variations in prescribing practices as instruments [4]; social epidemiology that propose geographic or temporal variation in policies as instruments [5]; and per-protocol analyses of randomized trials that propose random assignment as an instrument [6]. Each of these settings has unique challenges, but more often than not the lessons learned from or tools developed within one of these settings could be translated to the others. Given this, our discussion attempts to span all four settings.

For the purposes of this review, we always assume the goal is to obtain a numeric estimate for an average causal effect of a treatment or exposure on an outcome. Investigators using IV methods sometimes have other goals, including bounding causal effects or testing causal null hypotheses; reviews of these topics can be found elsewhere [7].

Condition (1): Relevance

The first and only verifiable condition is that the proposed instrument must be associated with the exposure. Verifying this only requires checking whether there is an association between the proposed instrument and the exposure. The proposed instrument does not need to cause the exposure but proxy instruments (i.e., instruments that are correlated with but do not cause the exposure themselves) can complicate the interpretation of effect estimates, as we will explain later [8,9].

Although the relevance condition (1) only requires an association exists, weak associations can mean that the analysis is vulnerable to weak instrument bias either via finite-sample limitations or by amplifying biases due to violations of other assumptions [10]. Therefore, strong instruments are generally preferred over weak instruments. However, deciding whether to use a proposed instrument (or choosing between multiple proposed instruments) solely based on strength, for instance using the F statistic or R^2 , can also lead to bias because these estimates are more likely to be overestimates of the true instrument strength

* The IV conditions have been formalized a number of ways, but for the purposes of this review the primary differences between these formalizations are not usually relevant. For more information on the various formalizations, see reference [7].

1
2
3
4 [11]. These measures of a proposed instrument's strength are also less directly relevant for understanding
5 whether and how a weak association between the proposed instrument and exposure would result in
6 amplifying biases due to violations of other assumptions.
7

8 **Condition (2): Exclusion restriction** 9

10 The exclusion restriction condition (2) requires that any effect of the proposed instrument on the outcome
11 is exclusively through its potential effect on exposure. This assumption is not verifiable. Therefore,
12 substantive knowledge of the relationship between the proposed instrument, the exposure and the
13 outcome must be used to justify the plausibility of the exclusion restriction.
14

15 There exist, however, methods of falsifying the exclusion restriction with the data, meaning that it is
16 sometimes possible to detect that the assumption is violated (but we cannot ever confirm that it holds).
17 More specifically, many of the available falsification tests jointly test condition (2) alongside condition
18 (3) discussed below. However, because there are settings where condition (3) is expected to hold by
19 design, these falsification strategies are sometimes described as applying to condition (2) alone, which
20 can aid interpretation of the reason for or degree of violation. For example, conditions (2)-(3) jointly can
21 be tested using the instrumental inequalities [12], which can be applied in many settings and can be easily
22 implemented as a one-sided test of a 2x2 table in the setting of a binary proposed instrument, binary
23 exposure, and binary outcome [13]. In the all-binary setting and assuming condition (3) holds, a detected
24 violation of the instrumental inequalities also provides some information on the subset of the study
25 population the violation of condition (2) occurs [14].
26
27
28

29 Other falsification strategies require leveraging additional substantive knowledge. When it is known, for
30 example, that a subgroup can be identified in which the proposed instrument does not affect the exposure,
31 any estimated association between proposed instrument and outcome must be due to a violation of
32 condition (2) or (3) [15,16]. For example, Kang and colleagues [16] check whether a genetic instrument
33 for malaria has an effect on their outcomes of interest in countries where malaria does not occur. In such
34 places, the genetic instrument cannot have an effect on malaria and any relationship with the outcome
35 must be due to a violation of condition (2) or (3). Assuming that the bias-causing mechanism is
36 homogeneous across subgroups and that the choice of subgroup does not induce selection bias, the bias
37 measured in this subgroup can be used to correct for the violation in the entire population [17]. Another
38 falsification strategy can be used if the confounding between the exposure and the outcome is known to
39 be positive (i.e., the confounded estimate is larger in magnitude than the true causal effect). This
40 knowledge implies specific relationships between the proposed instrument, exposure and outcome that
41 can be checked with the data [18].
42
43

44 **Condition (3): Exchangeability** 45

46 In essence, the exchangeability condition (3) takes the usual exchangeability assumption but forces us to
47 consider exchangeability for the proposed instrument rather than the exposure. Why then, if we are so
48 worried about the exchangeability assumption in traditional analyses, do we find it more plausible in IV
49 analyses? Proposed instruments are often selected because there are *a priori* reasons to believe that they
50 are exchangeable with respect to the outcome. For instance, if conducting an IV analysis in a randomized
51 trial or lottery study with random assignment as the proposed instrument, exchangeability is expected at
52 baseline by design. In other settings, this assumption must be argued to hold through subject matter
53 knowledge.
54
55

56 One indirect way of assessing this assumption is to look at whether there is imbalance in measured
57 covariates across levels of the proposed instrument, similar to covariate balance checks in randomized
58 controlled trials. Imbalance in measured covariates can in principle be eliminated by adjusting for them in
59 the analysis, but such imbalances can be suggestive of imbalances across unmeasured variables. As with
60
61
62
63
64
65

1
2
3
4 covariate balance checks in general, an understanding of the causal structure is needed to know whether
5 any perceived imbalances are potentially indicative of bias. In the causal diagram in Figure 1, X_1 , X_2 , X_3 ,
6 and X_4 are expected to be associated with the proposed instrument, but only imbalances in X_1 and X_2 reflect
7 violations of the instrumental conditions (conditions (3) and (2), respectively). Further, of course it is
8 possible that measured covariates appear balanced across levels of the proposed instrument but that there
9 is a lack of exchangeability due to unmeasured covariates: that is, again we see that we cannot verify
10 condition (3) but may sometimes find suggestive evidence against it. Finally, because the bias due to a
11 violation of condition (3) is a function of the proposed instrument's strength, it has been argued that these
12 types of covariate balance checks should be augmented to incorporate the proposed instrument's strength
13 into presentation, especially if presented in conjunction with a non-IV analysis [19].
14
15

16 Another indirect way of assessing this assumption is to look at negative outcomes, similar to negative
17 outcome controls used in non-IV studies [20,21]. Of course, such assessment requires the availability of a
18 secondary outcome that is not expected to be affected by the exposure (or the proposed instrument) but
19 could suffer from the same type of violation of condition (3) that the investigators are concerned about for
20 the primary research question.
21

22 **Estimating the average treatment effect with condition (4h): Homogeneity**

23
24 In order to obtain a point estimate, a fourth assumption is required and the choice of assumption
25 determines the causal parameter of interest. We first consider the condition (4h) under which the average
26 causal effect is identified.
27

28
29 The homogeneity assumption underlying the standard IV estimator requires that the proposed instrument
30 does not modify the effect of the exposure on the outcome among the exposed and unexposed on the
31 additive scale. If any unmeasured confounder of the exposure-outcome relationship is also an effect
32 measure modifier, then it is usually not reasonable to assume condition (4h) homogeneity [9]. Some
33 investigators propose conditioning on measured covariates that are perceived to be important effect
34 measure modifiers to recover the average causal effect [22,23].
35

36 Assessment of the homogeneity conditions remains difficult, and discussions of this condition remain
37 complicated in the literature because debates remain on how prevalent relevant heterogeneity is in
38 epidemiology [24]. (Of course, this varies depending on the study question.) Some relatively simple
39 checks have been proposed to falsify or understand the importance of condition (4h) in a particular study.
40 For example, conditions (1)-(3) alone allow for bounding of the average causal effect [12,25,26], which
41 means that when the bounds achieved under these three conditions alone are wide then it is at least
42 mathematically possible for the point estimate to be very different from the true causal effect due to a
43 violation of condition (4h).
44

45
46 In the simple setting of a dichotomous source of effect measure modification, Brookhart et al. [27]
47 showed that the bias due to a violation of condition (4h) is a function of how the strength of the proposed
48 instrument differs within strata of the modifier. Given this, the investigators proposed presenting how the
49 strength of the proposed instrument differs across measured covariates. The logic here is similar to that of
50 presenting covariate balance when considering condition (3): any detected differences in measured
51 covariates could theoretically be accounted for by including the modifier in the model, but may arguably
52 indicate that there are unmeasured sources of effect measure modification that violate condition (4h).
53
54

55 When both the proposed instrument and exposure are binary, and the proposed instrument causes the
56 exposure, then one can compare the counterfactual outcomes between "compliance types" as an indirect
57 assessment of condition (4h) [28]. A study participant's membership in one of the four mutually exclusive
58 compliance types is determined by how that person is affected by the instrument: people who are exposed
59 regardless of the instrument ("always-takers"), people who are never exposed regardless of the instrument
60
61
62
63
64
65

1
2
3
4 (“never-takers”), people who are coerced to be exposed because of the instrument (“compliers”) and
5 people who do the opposite of “compliers” (“defiers”). When there are no “defiers”, it is possible to
6 estimate the counterfactual outcome of both “never-takers” and “compliers” under no exposure from the
7 data. If these values are different, it is calls the homogeneity assumption into question. The same can be
8 done with the counterfactual outcomes of “always-takers” and “compliers” under exposure.
9

10
11 Finally, while the usual IV estimators tend to assume additive effect homogeneity, there is another IV
12 estimator based on a multiplicative rather than additive structural mean model that relies on a similar
13 assumption but on the multiplicative scale [9,25,29]. As homogeneity cannot simultaneously be satisfied
14 on both scales except under the null, careful thought should be given to whether one of these two scales
15 and therefore one of these two types of conditions (4h) is more likely to hold.
16

17 **Estimating the local average treatment effect with condition (4m): Monotonicity**

18

19 Economists, who are responsible for developing much of the early theory of IV methods, have generally
20 been skeptical of the homogeneity condition (4h) and instead looked toward a monotonicity condition
21 (4m) as a sometimes more plausible, alternative assumption. In the usual way that condition (4m) is
22 evoked, monotonicity requires that the proposed instrument only affects the exposure in one direction in
23 all individuals. In other words, there do not exist both people whose exposure level would have been
24 increased by increases in the proposed instrument and people whose exposure level would have been
25 decreased by increases in the proposed instrument. (Recently, different versions of monotonicity
26 conditions have been described that can change the interpretation of the effect estimate, but go beyond the
27 purposes of this review [30,31].) If monotonicity is assumed instead of homogeneity, an average causal
28 effect in the subgroup of “compliers” (described above) is identified [32]. Of note, some investigators
29 have argued against estimating this effect because the subgroup of “compliers” is not identified, and for
30 proxy or non-binary proposed instruments the interpretation of this subgroup becomes even less clear
31 [8,9,33]. The non-identifiability concern is partially mitigated when the proposed instrument and
32 exposure are binary and the proposed instrument is causal: then conditions (1)-(3) and (4m) allow us to
33 estimate the proportion of “compliers” as well as describe their characteristics in measured covariates
34 [34]. When the proposed instrument is a proxy (i.e., non-causal) instrument, estimating the proportion of
35 or characterizing the “compliers” becomes more difficult and requires additional assumptions and
36 considerations of the underlying causal instrument [8,9].
37
38
39

40 When the exposure is continuous, violations of condition (4m) can be found by graphing the difference in
41 cumulative distribution in exposure for each level of the proposed instrument [35]. If the difference in
42 cumulative distribution functions changes sign over the range of feasible exposure levels, then a violation
43 of the monotonicity assumption is detected. Failure to detect a violation, however, does not constitute
44 support for the monotonicity assumption.
45

46 When the proposed instrument is a measure of a decision-maker’s preference, such as the commonly
47 proposed provider preference instruments in pharmacoepidemiology studies, then it is also possible to
48 empirically assess the monotonicity condition (4m) by supplementing the data with a survey of the
49 providers [31]. By asking providers about their treatment decisions for the same set of (possibly
50 hypothetical) patients, investigators can measure the compliance type distribution and potentially find
51 evidence against monotonicity.
52
53

54 Finally, relatively simple inequalities can be checked to falsify monotonicity in the simple case of a
55 binary proposed causal instrument, binary exposure, and binary outcome (assuming conditions (1), (2),
56 and (3) hold). Coincidentally, these inequalities will be violated whenever the bounds on the average
57 causal effect proposed by Manski and Robins differ from those proposed by Balke and Pearl [12,25,26];
58 see reference [7] for more detail.
59
60
61
62
63
64
65

Settings with multiple proposed IVs

In some studies, investigators suggest that there is not just one but multiple proposed instruments. In such settings, it is sometimes possible to leverage this added potential information in ways that relax the instrumental conditions or that mitigate bias due to some of the proposed instruments not being instruments. For a more complete review of the available sensitivity analyses and robust methods, see reference [36]. We briefly highlight how having multiple proposed instruments can potentially address or detect some of the biases due to violations of the above-described conditions for each proposed instrument.

First, having multiple proposed instruments affords more ways to conduct an IV analysis. Investigators proposing multiple instruments have historically estimated causal effects by meta-analyzing individual estimates, by using all proposed instruments in the same two-stage least squares regression models, and by combining the proposed instruments into a summary risk score [37]. Generally, these methods can be more robust to weak instrument biases, although it is important to be aware of the assumptions made in each case. Newer estimators, primarily developed in the Mendelian randomization literature, can also allow for some types of violations of the exclusion restriction condition (2) [38–40] by leveraging additional assumptions (for a complete review of these additional assumptions, see reference [36]). MR-Egger in particular allows for certain violations of condition (2) by introducing homogeneity and linearity assumptions alongside the assumption that the strength of each proposed instrument is independent of the size of the direct effect violating condition (2); under these assumptions, MR-Egger can also be used to detect the existence of these types of violations of condition (2) for some of the proposed instruments [38].

Another oft-used falsification strategy in the multiple instrument setting is the Hausman over-identification test in which all proposed instruments are regressed on the residuals of a regression of the outcome on the exposure. It can identify whether at least one of the proposed instruments used is invalid but cannot identify which [41]. Moreover, it cannot detect which condition is violated among conditions (2), (3), or (4h). Such a test is especially underpowered and may not valid for weak instruments [42].

Importantly, because any sound method for estimating a causal effect or falsifying assumptions via multiple proposed instruments needs to be targeting the same average causal effect (rather than the instrument-specific causal effects within the “compliers”), these approaches all rely on a homogeneity condition (4h) or even stronger homogeneity or linearity conditions. This means that using multiple proposed instruments in settings where effect heterogeneity is likely (i.e., in which condition (4h) is unlikely to hold) makes interpreting effect estimates nearly impossible [43].

Beyond the core IV conditions

We have discussed the core assumptions that IV analyses make in replace of the usual exchangeability condition underlying non-IV analyses. However, there are some additional considerations worth noting. First, the core above-described assumptions replace our more typical assumption of no uncontrolled baseline confounding, but this means that any selection or information biases that could affect a non-IV analysis in the same study may also be problematic for an IV analysis. In addition to the more familiar selection biases (e.g., due to loss to follow-up [44,45]), subtler selection biases can arise in IV analyses that select on a subset of possible exposures [46,47]. Also, it is of course possible if not likely that in many observational studies with proposed instruments, the investigators may believe the above-described assumptions do not hold unconditionally but are more reasonable within levels of measured covariates; in such cases, investigators may consider applying the falsification strategies and tools described here within levels of the measured covariates.

1
2
3
4 Importantly, estimating an average causal effect in an IV or non-IV analysis alike requires having an
5 unambiguous definition of the intervention regardless of the estimation procedure used [48,49]. In some
6 studies employing IV analyses, such as with randomized trials or program evaluation, the intervention is
7 well-defined because it has been actually implemented. However, some studies employing IV analyses,
8 such as many Mendelian randomization studies, the intervention itself is not well-defined [50]. This
9 makes it very difficult to interpret or assess the validity of any presented effect estimates. Such vagueness
10 also often overlooks the fact that classical IV methods are developed in the context of time-fixed
11 treatments, and thus when the exposures can vary over time it is both less clear what investigators are
12 trying to estimate and why the above-described assumptions are reasonable [50–52] .
13
14

15 Finally, up until now, our consideration of falsification strategies and related tools has focused on
16 understanding the validity of an IV analysis. Sometimes IV analyses are performed alongside non-IV
17 analyses, and investigators are interested in understanding whether the IV analysis is more or less biased
18 than the non-IV analysis. Bias component plots have been proposed as one option for considering relative
19 bias due to violations in exchangeability across IV and non-IV methods [19]. Investigators also
20 sometimes begin by comparing the estimates from the two approaches, either using subjective criteria or a
21 formal test [41]. However, any detected differences could mean that the IV analysis is biased, the non-IV
22 analysis is biased, the analyses are estimating different causal effects (e.g., the effect in the “compliers”
23 vs. the average causal effect), or all of the above.
24
25

26 **Conclusion**

27
28 Addressing a causal question with methods that make different assumptions demonstrates whether
29 estimates are sensitive to the assumptions that differ across methods. In this way, estimates obtained from
30 IV analyses can nicely complement estimates from analyses that require measuring and appropriately
31 adjusting for confounders. However, the plausibility of these and any conditions required for causal
32 interpretations must still be verified when possible and, when verification is not possible, efforts must be
33 made to falsify the conditions as feasible. Here we have assembled a list of techniques that can be used to
34 falsify or weigh the reasonableness of the core conditions underlying IV analyses. By using all techniques
35 applicable to a specific epidemiologic IV analysis, we can use the data to its fullest extent.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Acknowledgements:

This work is supported by a NWO/ZonMW Veni grant [91617066].

Conflict of Interest

Jeremy Labrecque and Sonja A. Swanson declare no conflicts of interest.

Human and Animal Rights and Informed Consent

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29(4):722–9.
2. ••Swanson SA, Hernán MA. How to Report Instrumental Variable Analyses (Suggestions Welcome). *Epidemiology*. 2013;24(3):370–4. **Provides reporting guidelines for instrumental variable analyses.**
3. Boef AGC, Dekkers OM, Le Cessie S. Mendelian randomization studies: A review of the approaches used and the quality of reporting. *Int J Epidemiol*. 2015;44(2):496–511.
4. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the Reporting and Conduct of Instrumental Variable Studies. *Epidemiology*. 2013;24(3):363–9.
5. Glymour MM. Natural Experiments and Instrumental Variable Analyses in Social Epidemiology. In: Oakes MJ, Kaufman JS, editors. *Methods Soc Epidemiol*. 1st ed. San Francisco: Jossey-Bass; 2006. p. 429–56.
6. Swanson SA, Holme Ø, Løberg M, Kalager M, Bretthauer M, Hoff G, et al. Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening. *Trials*. 2015;16(1):541.
7. •Swanson SA, Hernán MA, Miller M, Robins JM, Richardson TS. Partial Identification of the Average Treatment Effect Using Instrumental Variables: Review of Methods for Binary Instruments, Treatments, and Outcomes. *J Am Stat Assoc*. 2018 (in press). **Provides a review of methods for bounding the average treatment effect under the different formal definitions of instrumental variables.**
8. Swanson SA, Hernán MA. The challenging interpretation of instrumental variable estimates under monotonicity. *Int J Epidemiol*. 2017;30(December):1–9 [Epub ahead of print].
9. Hernán MA, Robins JM. Instruments for causal inference: An epidemiologist’s dream? *Epidemiology*. 2006;17(4):360–72.
10. Nelson CR, Startz R. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *J Bus*. 1990;63(1):S125–40.
11. Hall AR, Rudebusch GD, Wilcox DW. Judging Instrument Relevance in Instrumental Variables Estimation. *Int Econ Rev (Philadelphia)*. 1996;37(2):283.
12. Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance. *J Am Stat Assoc*. 1997;92(439):1171–6.
13. Wang L, Robins JM, Richardson TS. On falsification of the binary instrumental variable model. *Biometrika*. 2017;104(1):229–36.
14. Richardson T, Robins J. Analysis of the binary instrumental variable model. In: Dechter R, Geffner H, Halpern JY, editors. *Heuristics, Probab Causality A Tribut to Jud Pearl*. London, England: College Publications; 2010. p. 415–44.

15. Altonji JG, Elder TE, Taber CR. An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling. *J Hum Resour.* 2005;40(4):791–821.
16. Kang H, Kreuels B, Adjei O, Krumkamp R, May J, Small DS. The causal effect of malaria on stunting: A Mendelian randomization and matching approach. *Int J Epidemiol.* 2013;42(5):1390–8.
17. •van Kippersluis H, Rietveld CA. Pleiotropy-robust Mendelian randomization. *Int J Epidemiol.* 2017;1–10 [Epub ahead of print]. **Describes a falsification strategy for the exclusion restriction.**
18. Glymour MM, Tchetgen EJT, Robins JM. Credible mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol.* 2012;175(4):332–9.
19. •Jackson JW, Swanson SA. Toward a Clearer Portrayal of Confounding Bias in Instrumental Variable Applications. *Epidemiology.* 2015;26(4):498–504. **Describes bias component plots.**
20. •Davies NM, Thomas KH, Taylor AE, Taylor GMJ, Martin RM, Munafò MR, et al. How to compare instrumental variable and conventional regression analyses using negative controls and bias plots. *Int J Epidemiol.* 2017;46(6):2067–77. **Describes negative controls in the context of instrumental variable analyses.**
21. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative Controls. *Epidemiology.* 2010;21(3):383–8.
22. Wang L, Tchetgen Tchetgen E. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J R Stat Soc Ser B (Statistical Methodol.)* 2017 [Epub ahead of print].
23. Ramsahai RR, Lauritzen SL. Likelihood analysis of the binary instrumental variable model. *Biometrika.* 2011;98(4):987–94.
24. Poole C, Shrier I, VanderWeele TJ. Is the Risk Difference Really a More Heterogeneous Measure? *Epidemiology.* 2015;26(5):714–8.
25. Robins JM. The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Heal. Serv. Res. Methodol. A Focus Aids.* 1989. p. 113–59.
26. Manski CF. Nonparametric Bounds on Treatment Effects. *Am Econ Rev.* 1990;80(2):319–23.
27. Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol.* 2009;169(3):273–84.
28. Abadie A. Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models. *J Am Stat Assoc.* 2002;97(457):284–92.
29. Hernán MA, Robins JM. Instruments for Causal Inference: An Epidemiologist’s Dream?: Erratum. *Epidemiology.* 2014;25(1):164.
30. Small DS, Tan Z, Ramsahai RR, Lorch SA, Brookhart MA. Instrumental variable estimation with a stochastic monotonicity assumption. *Stat Sci.* 2017;32(4):561-579.

- 1
- 2
- 3
- 4 31. Swanson SA, Miller M, Robins JM, Hernán MA. Definition and Evaluation of the Monotonicity
- 5 Condition for Preference-Based Instruments. *Epidemiology*. 2015;26(3):414–20.
- 6
- 7 32. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. *J*
- 8 *Am Stat Assoc*. 1996;91(434):444–55.
- 9
- 10 33. Deaton A. Instruments, Randomization, and Learning about Development. *J Econ Lit*. 2010;48:424–
- 11 55.
- 12
- 13 34. Angrist JD, Pischke JS. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, New
- 14 Jersey: Princeton University Press; 2009.
- 15
- 16 35. Angrist JD, Imbens GW. Two-Stage Least Squares Estimation of Average Causal Effects in Models
- 17 With Variable Treatment Intensit. *J Am Stat Assoc*. 1995;90(430):431–42.
- 18
- 19 36. ••Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian
- 20 randomization. *Stat Methods Med Res*. 2017;26(5):2333–55. **Reviews estimators used for Mendelian**
- 21 **randomization studies.**
- 22
- 23 37. Burgess S, Thompson SG. Avoiding bias from weak instruments in mendelian randomization studies.
- 24 *Int J Epidemiol*. 2011;40(3):755–64.
- 25
- 26 38. •Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect
- 27 estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512–25. **Describes**
- 28 **MR-Egger as a tool for both estimating effects and detecting bias in Mendelian randomization**
- 29 **studies.**
- 30
- 31 39. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian
- 32 Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*.
- 33 2016;40(4):304–14.
- 34
- 35 40. Tchetgen EJT, Sun B, Walter S. The GENIUS Approach to Robust Mendelian Randomization
- 36 Inference. *arXiv:1709.07779*. 2017;1–41.
- 37
- 38 41. Hausman JA. Specification Tests in Econometrics. *Source Econom*. 1978;46(46):1251–71.
- 39
- 40 42. Hahn J, Ham JC, Moon HR. The Hausman test and weak istruments. *J Econom*. 2011;160:289-299.
- 41
- 42 43. Swanson SA. Can We See the Forest for the IVs?: Mendelian Randomization Studies with Multiple
- 43 Genetic Variants. *Epidemiology*. 2017;28(1):43–6.
- 44
- 45 44. Boef AGC, le Cessie S, Dekkers OM. Mendelian randomization studies in the elderly. *Epidemiology*.
- 46 2015;26(2):e15-6.
- 47
- 48 45. Canan C, Lesko C, Lau B. Instrumental Variable Analyses and Selection Bias. *Epidemiology*.
- 49 2017;28(3):396–8.
- 50
- 51 46. Swanson SA, Robins JM, Miller M, Hernán MA. Selecting on treatment: A pervasive form of bias in
- 52 instrumental variable analyses. *Am J Epidemiol*. 2015;181(3):191–7.
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

47. Ertefaie A, Small D, Flory J, Hennessy S. A sensitivity analysis to assess bias due to selecting subjects based on treatment received. *Epidemiology*. 2016;27(2):e5–7.

48. Hernán MA, Robins JM. *Causal inference*. Boca Raton: Chapman & Hall/CRC; 2018.

49. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes*. 2008;32:S8–14.

50. • Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a Trialist? Deconstructing the analogy between Mendelian randomization studies and randomized trials. *Epidemiology*. 2017;28(5):653–9.

Leverages the analogy between Mendelian randomization studies and randomized trials to discuss strengths and limitations of Mendelian randomization as well as directions for future methodologic developments.

51. Swanson SA. Instrumental Variable Analyses in Pharmacoepidemiology: What Target Trials Do We Emulate? *Curr Epidemiol Reports*. *Current Epidemiology Reports*; 2017;4(4):281–7.

52. • Vanderweele TJ, Tchetgen EJT, Cornelis M, Kraft P. Methodological challenges in Mendelian randomization. *Epidemiology*. 2015;25(3):427–35. **Describes methodologic challenges in Mendelian randomization, underscoring several versions of violations of the exclusion restriction.**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 1: Causal directed acyclic graph of a proposed instrument Z , exposure A , outcome Y , and four additional covariates X_1 , X_2 , X_3 , and X_4 . By faithfulness, we would expect that Z would be associated with X_1 , X_2 , X_3 , and X_4 ; however, only associations with X_1 and X_2 indicate violations of the instrumental conditions. Additional unmeasured shared causes of variables in this graph are omitted to simplify presentation.

Table 1: Summary of falsification strategies and related tools for assessing the core conditions for an instrumental variable analysis.

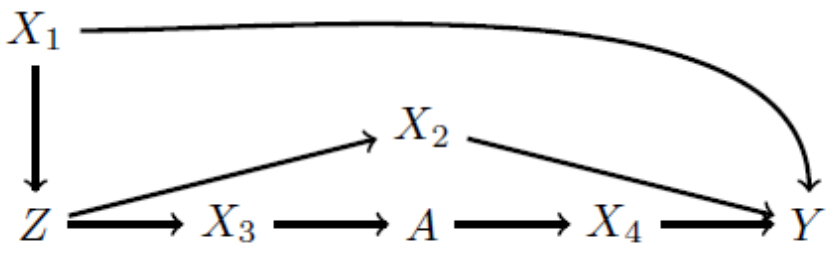


Table 1: Summary of falsification strategies and related tools for assessing the core conditions for an instrumental variable analysis.

Conditions	Strategy	Reference	Restrictions on the settings in which the strategy is applicable
(1)	Check association between instrument and exposure		N/A
(2), (3), (4h)	Over-identification	[41]	Multiple proposed instruments
(2), (3)	Leveraging positive confounding	[18]	Requires knowledge of the direction of confounding
(3)	Negative controls	[20]	Requires knowledge of the existence of an appropriate negative control
(2)	MR-Egger	[38]	Multiple proposed instruments; requires additional assumptions*
(2), (3)	Check in a subgroup where the instrument does not work	[15,16]	Requires knowledge of the existence of such a subgroup
(2), (3)	IV inequalities	[12]	Exposure cannot be continuous
(3)	Covariate balance and bias component plots	[19]	N/A
(4h)	Checking for differences in instrument strength across covariates	[27]	N/A
(4h)	Estimate counterfactual values among “always-takers”, “compliers” and “never-takers”	[28]	Condition (4m) must hold and the proposed instrument must be causal
(4m)	Cumulative distribution graphs	[35]	Exposure must be continuous
(4m)	Monotonicity inequalities	[12,25,26]	Causal binary proposed instrument, binary exposure, binary outcome
(4m)	Survey of provider preferences	[31]	Proposed instrument must be preference

*See text for further description of the additional assumptions.