SportRXiv

Part of the Society for Transparency, Openness and Replication in Kinesiology (STORK)

Preprint
not peer reviewed

Received: 31 December 2020
For correspondence:
ian.shrier@mcgill.ca

Methods Matter: Some important challenges with instrumental variable methods.

Ian Shrier[1], Tyrel Stokes[2], and Russell J. Steele[2]

[1] Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal Canada, [2] Department of Mathematics and Statistics, McGill University, Montreal, Canada

Please cite as: Shrier, I., Stokes T, & Steele RJ. (2020). Methods Matter: Some important challenges with instrumental variable methods. SportRχiv. doi: 10.31236/osf.io/ve4na

Address Correspondence to:
Ian Shrier MD, PhD
Centre for Clinical Epidemiology
Lady Davis Institute, Jewish General Hospital
3755 Cote Sainte Catherine Road
Montreal, QC H3T 1E2, Canada
Tel: 514-340-7563
Email: ian.shrier@mcgill.ca

All authors have read and approved this version of the manuscript.
This article was last modified on January 6, 2021.

ABSTRACT

A recent editorial in the British Journal of Sports Medicine (BJSM) suggested instrumental variable (IV) analysis has advantages in estimating causal effects when there is low compliance. We originally submitted a version of this commentary to BJSM as an editorial (they do not have a letter to editor section) but it was rejected without review. The original BJSM editorial included several important errors, presented results that are inconsistent with the results of an IV analysis, and omitted definitions and important limitations. All of these factors contributed to inappropriate interpretations. This commentary highlights the most important errors. We also believe the BJSM editorial serves as another reminder that appropriate statisticians should be included from the beginning of the study wherever possible. At the very least, they should be the co-authors responsible for calculating results and ensuring the write-up is consistent with the results.

INTRODUCTION

A recent editorial in the British Journal of Sports Medicine (BJSM) suggested instrumental variable (IV) analysis has advantages in estimating causal effects when there is low compliance.[1] We agree that IV analyses can be valuable in addressing certain questions when there is unmeasured confounding.[2 3] However, the BJSM editorial presented results that are inconsistent with the results of an IV analysis, and omitted definitions and important limitations. We are concerned this will lead to important misinterpretations and inappropriate decisions regarding the effects of treatment or prevention programs.

We have two objectives in this article. First, we want to make sure readers correctly understand what instrumental variable analyses can and cannot do. We originally submitted a version of this article to BJSM as an editorial (they do not have a letter to editor section) that commented only on the major issues of the previous editorial.[1] but it was rejected without review.

Second, and more importantly, we have recently argued that appropriate methodologists and statisticians need to be included when developing consensus statements on methods in sport and exercise medicine research.[4] We believe the errors in the BJSM editorial illustrate why similar recommendations should be applied to almost every analytical research paper. Each paper should include a statistician as a co-author (there was none on this paper) *who assumes responsibility for conducting and reporting of analyses*.

Most sport medicine clinicians would be surprised if a group of statisticians were capable of authoring a meaningful paper on sport medicine injuries without being specifically trained in the substantive material. Much like epidemiology and sports medicine, statistics is a broad, nuanced, and sometimes complex field requiring years of specific training and expertise to carry out and evaluate properly. This is particularly true when evaluating the importance and consequences of assumptions and modelling choices in complex domains or with advanced methodologies such as repeated measures on the same

DOI: 10.31236/osf.io/ve4na

participant, recurrent injuries, and modern causal estimators. If we believe that "Methods Matter", let us ensure that we conduct and report our research appropriately.

MAKING SENSE OF TABLE 1

First, Table 1 of the original editorial (reproduced in our Table 1) contained elementary errors. Some of these errors were due to transcription according to the senior author (personal communication). For example, the point estimate for the "risk in intervention group" under Intention to Treat analysis is 66.8%, which is outside the reported 95% confidence interval (95%CI) of -7% to 12.5%. Second, the lower risk of -7% is not possible because a risk cannot be lower than 0 (or greater than 1). There is a similar problem under the Instrumental Variable analysis, where the point estimate for risk in the "compliance" group is -141.9%. This type of result can occur if the 95%CI is calculated as estimate ± 1.96*SE. However, when this occurs, statisticians recommend other methods to avoid presenting results that are not possible.[5]

Table 1. Reported results from [1]

	Risk/Risk difference % (95% CI)
Intention-to-treat analysis	
Risk in control group (ref)	69.9 (62.2 to 77.6)
Risk in intervention group	66.8 (-7.0 to 12.5)
Cumulative risk difference	-3.1 (-12.9 to 6.6)
As treated analysis	
Risk in non-compliance group (ref)	67.8 (63.0 to 72.7)
Risk in compliance group	80.7 (53.8 to 107.6
Cumulative risk difference	12.9 (-14.4 to 40.3)
Instrumental variable analysis	
Risk in non-compliance group (ref)	70.2 (61.7% to 78.6%)
Risk in compliance group	21.7 (−141.9% to 185.4%)
Cumulative risk difference	−52.5 (−218.7% to 113.7%)

These Table 1 numerical errors are a minor nuisance because we expect the authors to publish an erratum to correct them, and the erratum will be indexed in search engines. However, Table 1 includes other errors that could lead to serious misunderstandings.

1. In the As Treated analysis, the authors reported risk in the "non-compliance" and "compliance" groups. In the text, the authors appear to define compliance groups as those who are observed to follow treatment assignment: "… 95% of the compliant athletes reported a history of injury in the previous season." However, according to their definition, "compliant athletes" include both athletes who are assigned control treatment and take control treatment as well as athletes who

are assigned active treatment and take active treatment. If the "compliance" label were correct, the estimate in the "compliance" group would represent a mixture of athletes who were treated and untreated. The same problem occurs for the "non-compliant" group. Therefore, comparing "compliance" to "non-compliance" groups does not represent an As Treated analysis. The authors graciously provided the statistical code. From the code, we believe the label for non-compliance group should be "Risk in group that did not receive active treatment" and the label for the compliance group should be "Risk in group that did receive active treatment", where active treatment is defined as receiving an average of >2 training sessions per week over the entire study period (see footnote[1]).

2. The Instrumental Variable analysis section presents a more serious challenge. In the context of adherence within an RCT, standard IV analysis provides the equivalent of a risk ratio (hazard ratio in the BJSM editorial because they used a time-to-event analysis) between "compliers" and "non-compliers", so the labels could be correct. However, there are two issues. First, the IV method defines compliers differently than the authors did, and differently from the As Treated analysis above (see below). In addition, the analysis can only give the risk ratio; it cannot give absolute numbers for each group, although they can be obtained with other methods.[3] Authors should report results as per published recommendations.[6]

Beyond Table 1, the manuscript text includes many statements that are likely to be misinterpreted and may lead to inappropriate inferences if repeated in future studies.

BIAS IS RELATED TO THE RESEARCH QUESTION

In the text, the authors also make broad statements about bias without recognizing that the existence of bias depends on the particular research question. More specifically, we make the following points about the authors' discussion of bias under different analyses.

1. The authors imply the intention-to-treat (ITT) analysis is always biased when compliance is low (Figure 1, next page).[1] "Bias" can only be interpreted in response to a specific question, which was not stated in the editorial. For the public health question: "What is the population average causal effect of assigning treatment?", the results of the ITT analysis are generally unbiased whereas the results of the IV analysis are generally biased.[3]
2. We believe the authors were asking "What is the population average causal effect of treatment (PACE) when treatment is taken as prescribed?" We agree the results of the ITT, "As Treated"

[1] **Footnote:** Defining "observed compliance" is not always straightforward and needs to be reported clearly. Readers should be aware that in the published analysis, participants in the assigned active treatment group "were asked to perform the specifically designed AIPP twice a week." Therefore, a participant who followed the recommended program (=2x/week) would not be considered as receiving active treatment in the analysis (required average >2x/week), and a participant who alternated between training 5x/week and 0x/week (average>2x/week) would be considered as receiving active treatment.

DOI: 10.31236/osf.io/ve4na

and "Per Protocol" analyses are generally biased for this question.[3 7] However, the results of the IV analysis are also biased except under very restricted conditions.[8]

3. Although the results of the IV analysis are biased for PACE, they are unbiased for the average causal effect in a subpopulation known as "compliers" (complier average causal effect, CACE) that is different from the authors' definition of compliers. For clarity, we refer to the authors' definition as "observed compliers". This is elaborated on in the next section.

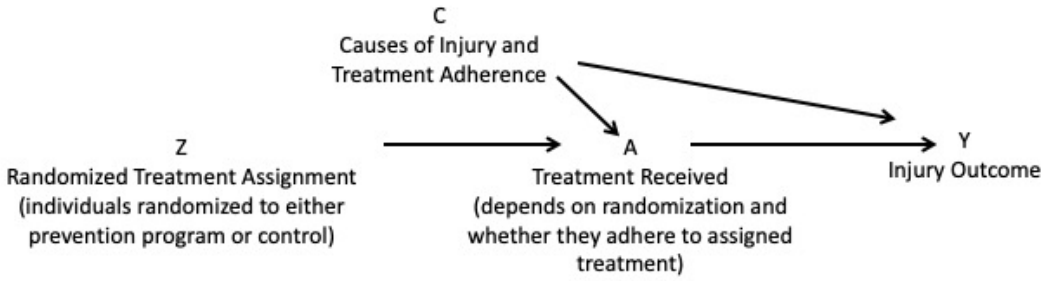

Figure 1. A causal directed acyclic graph (DAG) for a randomized trial. Some individuals do not adhere to their assigned treatment (Z) because of confounding factors (C) that also cause the outcome (Y). Treatment received (A) depends on both randomization status and the confounding factors. The outcome depends on the treatment received as well as the confounding factors. There are several different instrumental variable methods to estimate the complier average causal effect (CACE). As one example, we can easily go from the estimate for the ITT (intention to treat effect) to the CACE by the simple formula:
$CACE = \frac{ITT}{P(Complier)}$ (see footnote[2]).[9]
Note that this analysis does not yield separate estimates of risk for those who received treatment or did not receive treatment. Other IV methods for estimating CACE have the same limitation. It is not clear how the authors obtained the estimates they reported for these different groups under the Instrumental Variable analysis in Table 1.

[2] **Footnote**: When we make the usual assumptions required for causal inference[2], the causal effect of assigning treatment on the injury outcome for each participant (i.e the Intention to Treat (ITT) estimate) is the product of the causal effect of assigning treatment on receiving treatment for the participant, and the causal effect of receiving treatment on the outcome for the participant. The population average causal effect (PACE) is then a weighted average of the causal effect within the different strata, i.e. always takers, never takers, IV compliers and defiers. described in the section "Instrumental Variable "Compliers" vs "Observed Compliers". Mathematically, the only strata that contribute to the average with IV methods are those that change their behaviour based on assignment, i.e the compliers and defiers, which mathematically can be written as $A_i(Z_i = 1) \neq A_i(Z_i = 0)$. When we assume no defiers, the ITT is equal to the average causal effect in the compliers (i.e. CACE: $E[Y_i(A_i = 1) - Y_i(A_i = 0)|A_i(Z_i = 1) - A_i(Z_i = 0) = 1]$) multiplied by the probability of adhering to the treatment assignment ($P(A_i(Z_i = 1) - A_i(Z_i = 0) = 1)$), i.e. $ITT = CACE * P(Complier)$. If the causal effect is different in always takers and never takers compared to compliers and defiers, the CACE will be different from the PACE. Also, when the probability of adhering to treatment assignment is 1 (100% adherence), the ITT, CACE and PACE are all equal. As this probability becomes small we can see that the ITT result will become increasingly biased for both the CACE and PACE. However, the ITT is always unbiased for the causal effect of assigning treatment in this set-up.

DOI: 10.31236/osf.io/ve4na

INSTRUMENTAL VARIABLE "COMPLIERS" VS "OBSERVED COMPLIERS"

Randomized trial participants can be divided into four groups prior to the trial based on what their behaviour *would be* if assigned control or active treatment[3 8 10]:

1. "Compliers" are those who would always follow their assigned treatment regardless of group assignment
2. "Always takers" are those who would *always* take the active treatment regardless of group assignment
3. "Never takers" are those who would *never* take the active treatment regardless of group assignment
4. "Defiers" are those who would always take the opposite of their assigned treatment (usually assumed not to exist)

The authors' "observed compliers" refer to participants who were assigned a particular treatment and received that treatment. If "defiers" do not exist, participants assigned active treatment who did not take active treatment must be never takers. Because participants are randomized, we expect the assigned control group to have an equal number of never takers as the assigned active treatment group. These never takers in the control group are considered "observed compliers" according to the original editorial's description because they were assigned control and received control. Therefore, observed compliers in the control group are a mix of never takers and IV compliers (those that would have taken treatment if assigned treatment).[3 10] Similarly, when active treatment is available outside the trial, some participants in the assigned control group may receive active treatment, and are therefore always takers. As with never takers, we expect an equal number of always takers in the assigned active treatment group as occurred in the assigned control group. Therefore; the observed compliers in the assigned active treatment group are a mix of always takers and IV compliers.

The estimate from an IV analysis only applies to the subpopulation of IV compliers, not observed compliers. This is commonly referred to as complier average causal effect (CACE) or local average treatment effect (LATE).[8] When the effect in IV compliers varies across compliers, always takers and never takers, the results of the IV analysis are unbiased for CACE, but biased for PACE. Further, in the BJSM editorial, the CACE applies to a maximum of 6.9% of their study population (i.e. the observed compliers). In their study, a few control participants received active treatment (personal communication with senior author). These participants were always takers, and therefore the proportion of IV compliers is slightly less than the 6.9% reported.[2]

From the above point, it becomes obvious that the instrumental variable methods address a different question than other analyses; they do not overcome limitations due to "low compliance" as implied by the title or conclusion of the BJSM editorial.

SOME IMPORTANT ASSUMPTIONS WITH IV METHODS

In this section, we highlight what we consider the two most important assumptions that were not included by the authors for unbiased IV analysis results.

1. One must assume defiers do not exist in the study, or use advanced sensitivity analyses with additional assumptions.[8] Although plausible in many trials, it needs to be evaluated.[8]

2. Assigning treatment cannot affect the outcome except through treatment.[6] For example, if participants assigned an injury prevention treatment decide to improve on the program by adding additional stretching / strengthening exercises (that affect injury rates), this assumption would be violated.

OTHER CHALLENGES WITH IV METHODS

We agree with the authors and have written that IV analyses might be useful in sport and exercise medicine.[2 3] However, readers should be aware that several influential researchers in causal inference believe its limitations often outweigh its usefulness. [8 11] For example, we cannot identify if patients who come to see us are IV compliers, always takers or never takers.[8] Therefore, simply citing the IV result provided would be misleading (biased) if one wants to know the PACE. We have previously argued that these types of analyses always be accompanied by sensitivity analyses and provided methods to do so.[12]

The BJSM editorial commented on the very wide confidence intervals they observed. In general, IV methods usually require very large sample sizes, and are usually conducted with data from large databases. They are unlikely to be useful for small studies, and the problems are worse when the outcome is categorized as Yes/No (e.g. injury), as opposed to continuous outcomes.

We also want to highlight some challenges with changing terminology. The PACE for taking treatment is referred to by some researchers as the per protocol effect [7]. Although the common "per protocol analysis" is usually biased for PACE, the PACE can be estimated using g-methods,[7] again providing the underlying assumptions (different from IV methods) are true.

SUMMARY

We believe that IV methods represent an important potential tool for sport and exercise medicine investigators. However, authors need to properly apply methods, understand and clearly describe their results, and the interpretations need to be consistent with the underlying principles of the analysis. The BJSM editorial serves as another reminder that experienced sport medicine clinicians and investigators should collaborate closely with the appropriate statistical expertise. Appropriate statisticians should be included from the beginning of the study wherever possible. At the very least, they should be the co-authors responsible for calculating results and ensuring the write-up is consistent with the results. Sport and exercise medicine research is important because it can help reduce injuries, illness and improve the quality of life of our patients. It is time we insist on incorporating processes that minimize the risk of making inappropriate inferences so that we are more likely to recommend appropriate treatment and improve the quality of life of our patients and athletes.

DOI: 10.31236/osf.io/ve4na

CONTRIBUTIONS

Contributed to conception and design: IS, TS, RJS
Drafted and/or revised the article: IS, TS, RJS
Approved the submitted version for publication: IS, TS, RJS

ACKNOWLEDGEMENTS

None

FUNDING INFORMATION

This work was unfunded

DATA AND SUPPLEMENTARY MATERIAL ACCESSIBILITY

Not applicable

REFERENCES

1. Edouard P, Steffen K, Navarro L et al. Methods matter: dealing with low compliance in sports injury trials analyses using instrumental variable analysis. *Br J Sports Med* 2020.
2. Wang C, Vargas JT, Stokes T et al. Analyzing activity and injury: Lessons learned from the acute:chronic workload ratio. *Sports Med* 2020;50:1243-1254.
3. Shrier I, Steele RJ, Verhagen E et al. Beyond intention-to-treat: what is the right question? *Clin Trials* 2014;11:28-37.
4. Shrier I. Consensus statements that fail to recognise dissent are flawed by design: a narrative review with 10 suggested improvements (In Press). *British Journal of Sports Medicine* 2020.
5. De Angelis D, Young GA. Bootstrap method. In: Armitage P, Colton T, editors. *Encyclopedia of Biostatistics*. 2nd ed: Wiley, 2005:DOI: 10.1002/0470011815.b2a14005.
6. Swanson SA, Hernan MA. How to report instrumental variable analyses (Suggestions welcome). *Epidemiology* 2013;24:370-374.
7. Hernan MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *N Engl J Med* 2017;377(14):1391-1398.
8. Swanson SA, Hernan MA. The challenging interpretation of instrumental variable estimates under monotonicity. *Int J Epidemiol* 2017.
9. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables *J Amer Statist Assoc* 1996;91(434):444-455.
10. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;58(1):21-29.
11. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006;17(4):360-372.
12. Steele RJ, Shrier I, Kaufman JS et al. Simple estimation of patient-oriented effects from randomized trials: an open and shut CACE. *Am J Epidemiol* 2015;182:557-566.