**Collider bias from selecting disease samples distorts causal inferences**

Gibran Hemani1,\*, Kate Tilling1, George Davey Smith1

1. MRC Integrative Epidemiology Unit, University of Bristol

Corresponding author: g.hemani@bristol.ac.uk

Dear Editor

A recent paper appeared in Genetic Epidemiology alerting readers to issues pertaining to a sensitivity analysis used in Mendelian randomisation (MR) analyses known as the MR Steiger directionality test1. In our original paper describing MR Steiger2 we warned readers that unmeasured confounding, horizontal pleiotropy and measurement error could all bias the MR Steiger result, and supported this with theory and simulations. We provided a sensitivity analysis in our software for measurement error3, and Lutz et al have added further sensitivity analyses for horizontal pleiotropy in a separate R package reverseDirection. We do however feel it necessary to draw attention to some errors or misinterpretations in their paper.

Lutz et al present an empirical analysis of an MR analysis of smoking cessation (former vs current smoking) on lung function (Forced expiratory volume in the first second, FEV1) and report that the MR Steiger method infers the wrong causal direction. They attribute this incorrect result to the method’s sensitivity to horizontal pleiotropy. We offer an alternative explanation in this case. The study selected only individuals who had ever smoked and who were ascertained for chronic obstructive pulmonary disease (COPD), and this can give rise to two collider paths (**Figure 1**) which can distort the R2 estimates compared to the general population:

1. if COPD is influenced by both smoking behaviour and FEV1 then the relationship between genetic instruments and lung function will likely become distorted in the selected sample due to collider bias. This will distort both the MR analysis and can also distort the MR Steiger result.
2. If smoking influences FEV1 then sample selection by restricting to ever smokers opens a collider path from the instrument to FEV1 via the unmeasured confounder U, and so depending on the relative magnitude of confounding on smoking and FEV1 the instrument-outcome R2 could be estimated to be higher than the instrument-exposure R2.

See the **Supplementary Note** for a numerical example that illustrates both of these mechanisms. We would not describe these as issues with the MR Steiger test, as the incorrect result is a consequence of bias in estimating the R2 values due to the study design.

We re-performed the analysis that Lutz et al present, this time using publicly available UK Biobank data, which does not suffer from sample selection due to COPD status (though important to note it does suffer from other forms of less extreme sample selection4). In our analyses cigarette pack years and ever vs never smoking have negative influences on lung function, while smoking cessation (which selects for ever smokers) has a positive influence on lung function. In all analyses smoking behaviour are inferred to be causally upstream of lung function using the MR Steiger test (**Table 1**, full methods and results in **Supplementary Note**). This result contradicts Lutz et al’s claim that horizontal pleiotropy distorted the results of the MR Steiger method, as we would expect horizontal pleiotropic effects to be relatively consistent across datasets when using the same instruments and when analysing individuals from the same ancestral background. To our knowledge Lutz et al did not make the summary data available for the analyses that they ran on the COPD selected sample, thus we cannot re-analyse that directly.

Lutz et al state that they performed the MR Steiger directionality test for the orientation of the exposure on the outcome, which would contrast the total instrument-exposure R2 against the total instrument-outcome R2 value to determine the causal direction between the traits. However, in their analysis there are multiple instruments available for the exposure and they present MR Steiger results separately for each instrument. We have previously described the per-instrument approach as ‘Steiger filtering’5. It is intended to be used in the case of many independent instruments being available for an MR analysis, to detect whether any may be invalid due to primarily associating with the outcome in a reverse-directional relationship6, or through horizontal pleiotropy. The per-instrument Steiger filtering is not sensitive to pleiotropy as implicated by Lutz et al, rather, it is a test of pleiotropy. The list of caveats of winner’s curse, measurement error and unmeasured confounding apply to Steiger filtering as much as they do to the MR Steiger directionality test. Nevertheless we performed Steiger filtering in the smoking cessation - FEV1 analysis using the publicly available data which has not been selected for COPD, and using the same instruments as those selected by Lutz et al. We show that none of these instruments are flagged as reverse causal, which is in contradiction to their results (**Table 1**).

On a separate note, Lutz et al mention that unmeasured confounding does not invalidate the MR Steiger test. This is incorrect; unmeasured confounding does remain an issue in some circumstances. The MR Steiger test is based on the comparison of R2 values, in which causal direction is inferred by determining whether the instrument-exposure R2 is larger than the instrument-outcome R2. When the net confounding effects influence the exposure and outcome in opposite directions and the causal effect of exposure on outcome is substantively large then it is possible that the instrument-outcome R2 value can exceed the instrument-exposure R2 value even when the true causal direction is from exposure to outcome. Theory and simulations in the original paper (sections S3 Text and S2 Figure) illustrate this problem. Lutz et al made the mistake of only simulating data where the directions of confounding and causal effect are all concordant – in which case, the MR Steiger test will identify the correct causal direction irrespective of the size of the confounding.

## Acknowledgements

GH is funded by the Wellcome Trust and Royal Society [208806/Z/17/Z]. KT and GDS work in the Integrative Epidemiology Unit which receives funding from the U.K. Medical Research Council and the University of Bristol (MC\_UU\_00011/3 and MC\_UU\_00011/1).

## Data availability

All data is made available via the OpenGWAS database <https://gwas.mrcieu.ac.uk/>

All code used to perform the analysis is available at <https://github.com/explodecomputer/steiger_collider/>

## References

1. Lutz, S. M., Wu, A. C., Hokanson, J. E., Vansteelandt, S. & Lange, C. Caution against examining the role of reverse causality in Mendelian Randomization. *Genet. Epidemiol.* **45**, 445–454 (2021).

2. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genet.* **13**, e1007081 (2017).

3. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, (2018).

4. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* **47**, 226–235 (2018).

5. Hemani, G. *et al.* Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv* (2017).

6. Anderson, E. L. *et al.* Education, intelligence and Alzheimer’s disease: evidence from a multivariable two-sample Mendelian randomization study. *Int. J. Epidemiol.* **49**, 1163–1172 (2020).

[Graphical user interface, application

Description automatically generated](https://app.diagrams.net/?page-id=uXIm_7eZGZ7xztSpjhyR&scale=auto#G19WZcqCQA2fe2j5Byq-PD72ONXUR5l_YP)

**Figure 1: A DAG depicting the analytical design in Lutz et al 2021.** The model being analysed estimates the causal influence of smoking status on lung function, however it is performed in a sample that has been selected for individuals who have ever smoked and who have COPD. Selecting samples in this manner can induce collider bias, in two ways. First, when selecting for ever-smokers, there is a collider path opened from the instrument (G) to the confounder (U). Second, when selecting for COPD there is a collider path opened from G to the outcome, lung function. Generally we would recommend avoiding performing any analysis that induces such complex collider issues because any inference will be difficult to interpret if the analysis is attempting to understand causal relationships in the general population.

**Table 1: MR and MR Steiger results for analysis of smoking behaviours on forced expiratory volume**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Exposure | # SNP | MR estimate | MR SE | R2(g,x) | R2(g,y) | Correct Steiger | Steiger p-value |
| Ever vs never smoked, UK Biobank | 84 | -0.066 | 0.060 | 0.0072 | 8.04E-04 | TRUE | 2.30E-157 |
| Cigarette smoked per day, Liu et al 2019 | 23 | -0.063 | 0.014 | 0.0089 | 4.38E-04 | TRUE | 1.86E-189 |
| Smoking cessation, Furberg et al 2010, using all Lutz et al selected instruments | 5 | 0.150 | 0.029 | 0.0019 | 3.96E-04 | TRUE | 3.15E-06 |
| Smoking cessation, UK Biobank | 1 | 0.471 | 0.122 | 0.0004 | 3.51E-05 | TRUE | 2.50E-05 |
| Smoking cessation, rs11633958 | 1 | 0.161 | 0.025 | 3.74E-04 | 6.11E-05 | TRUE | 0.106 |
| Smoking cessation, rs2869548 | 1 | 0.165 | 0.029 | 2.88E-04 | 4.96E-05 | TRUE | 0.164 |
| Smoking cessation, rs7260329 | 1 | -0.102 | 0.057 | 7.44E-05 | 4.90E-06 | TRUE | 0.368 |
| Smoking cessation, rs72738786 | 1 | 0.171 | 0.025 | 3.77E-04 | 7.03E-05 | TRUE | 0.121 |
| Smoking cessation, rs8192482 | 1 | 0.155 | 0.050 | 3.82E-04 | 5.81E-05 | TRUE | 0.094 |

MR estimates are from the inverse variance weighted method unless there is only one SNP in which case it is the Wald ratio estimate.

**Supplementary Note:** <https://explodecomputer.github.io/steiger_collider/lutz_reanalysis.nb.html>