

Can't see the wood for the trees: confounders, colliders and causal inference - a statistician's approach

Bin Huang,^{1,2} Rhonda Szczesniak^{1,2}

In a recent publication, Lederer and 47 other editors, including contributors from *Thorax*, put forth firm guidelines to improve the rigour and reproducibility of causal inference studies using observational data in respiratory research.¹ The authors highlight a continued reliance on antiquated methods and use of inappropriate procedures to account for confounding as reasons for a call to action. So, why and how must we carefully control for confounding? The journal includes an editorial offering the clinician perspective.² We complement this editorial from a statistical standpoint.

Concern over confounding has persisted throughout the history of observational studies. In clinical effectiveness studies, unadjusted confounding can lead to 'unfair' comparisons. A well-known instance of confounding is referred to as Simpson's paradox.³ Let's take a look at a hypothetical example presented in table 1, comparing the performance of two treatments (labelled A and B) and their success rates for treating pulmonary exacerbations in cystic fibrosis patients. Here, successful treatment for each patient is defined as returning to a pre-exacerbation level of forced expiratory volume in FEV1% predicted. If you were only given the average rate of success between A and B, as shown by the overall success rates, it may lead you to believe that treatment A is doing a better job at combatting pulmonary exacerbations than treatment B. However, if you were provided the breakdown of the patient's condition (mild or severe), it becomes quite obvious that treatment B is consistently out-performing treatment A, regardless of condition. So, what's going on? Confounding is playing a trick on us! The underlying disease severity confounds the relationship between treatments and their success rates. Treatment B, which is believed to be more effective,

has a 80% chance being prescribed to severe patients and successfully helps 62.5% of them. Meanwhile, only 20% of severe patients were prescribed treatment A, yielding a success rate of only one in two severe patients. However, the overall summary aggregates all patients and the success rate despite underlying disease severity, thereby unfairly lowering the overall success rate for treatment B. This example demonstrates clearly why we should account for confounding and highlights the danger of ignoring it. So, how do we account for confounding, in order to know what treatment is truly better?

The authors thoroughly outline principles and recommendations for recognising and accounting for confounding, and they tackle several challenges with confounding in observational studies. They use a simplified example to illustrate how smoking confounds the association between exercise and lung cancer. They make an important suggestion to alleviate confounding through use of a directed acyclic graph (DAG), which can model various kinds of information on the causal underpinnings of the research question. Explicitly depicting the underlying prior beliefs in an observational setting is critically important. Without a DAG, the door for misinterpretation and drawing misleading conclusions is left wide open.

For illustration of the DAG, let's consider an obvious example in health disparities research using observational data. Researcher A postulates that poor health outcome (denoted as Y) in racial minorities (exposure denoted as X) is mediated through lower family socioeconomic status (SES) as shown in figure 1.

Researcher B, however, believes SES confounds—rather than mediates—race and health outcome linkage as depicted in figure 2. In addition, both researchers acknowledge the complex relationships and existence of unmeasured confounders (denoted as U), such as neighbourhood and exposure to stressors (denoted as SES $\leftarrow U \rightarrow Y$). Under the DAG in figure 1 (a mediation model), it is clear that the total effect of race on Y is not identifiable. While in the DAG in figure 2 (a confounding model), both the total and direct effects between race and outcome Y can be identified, as long as we have adjusted for the confounder SES. Without the DAG's explicit depiction, researcher A may proceed with regression modelling, including SES in the equation, and claim that SES mediates the racial disparity. Yet in fact, the results only support association rather than mediation. As pointed out by the guideline authors, very few observational research studies present DAGs and discuss the impact that unmeasured confounding can have on study results. While researchers may not always agree on the causal underpinnings and the DAG cannot itself ameliorate sources of confounding, the explicit presentation of the DAG facilitates discussion of these differences and promotes understanding of how distinct conclusions were reached. Encouraging the use of DAGs may further ensure the reproducibility and integrity necessary to establish causal relationships.

To some researchers, these guidelines may provoke questions about the validity of observational studies. For example, why not conduct a randomised controlled trial (RCT) to answer causal questions? Well known as the 'gold standard' for research and hypothesis testing, RCTs draw on targeted populations using highly controlled protocols. On the other hand, observational studies typically rely on a heterogeneous population examined in a broad range of settings. For these reasons, should we compromise internal validity for external validity with an observational study? To consider this

Table 1 Performance of treatments, overall and by condition

Treatment	Condition	Outcome		Success rate
		Success	Failure	
A	Mild	175	25	175/200=87.5%
	Severe	25	25	25/50=50%
	Overall	200	50	200/250=80%
B	Mild	45	5	45/50=90%
	Severe	125	75	125/200=62.5%
	Overall	170	80	170/250=68%

¹Division of Biostatistics & Epidemiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
²Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA

Correspondence to Dr Rhonda Szczesniak, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati OH 45229, USA; Rhonda.Szczesniak@cchmc.org

Mediation Model

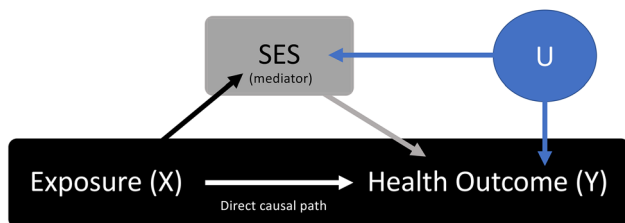


Figure 1 Mediation model as shown using a directed acyclic graph. SES, socioeconomic status.

question, let's get back to the Simpson's paradox example in table 1. In this hypothetical scenario, suppose we conduct an RCT to find out which treatment truly works better. We may recruit and randomise 240 patients to be treated by therapy A and 240 to be treated by therapy B. After many considerations, such as the inclusion and exclusion criteria, as well as patient's willingness to be randomised into each arm, the trial includes 200 mild and 40 severe patients. Then, the estimated success rate in the 240 patients randomised to treatment B is $200 \times 90\% + 40 \times 62.5\% = 205/240$. On the other hand, the success rate in 240 patients randomised to treatment A is $200 \times 87.5\% + 40 \times 50\% = 195/240$. Therefore, the trial rightfully declares that treatment B is better and concludes that 10/240 or 4.2 cases out of 100 patients is attributable to better performance of treatment B. By randomisation, the patient populations assigned to the treatment arms are completely exchangeable. Thus, the internal validity is ensured. However, suppose the table 1 tabulates the distribution of the entire patient populations; therefore, the true distribution of the patient population is 50% in the mild condition and 50% in the severe condition. Then, $(90 - 87.5) \times 50\% + (62.5 - 50) \times 50\% = 7.5\%$, or out of every 200 patients 15 additional successes would have been achieved if they were treated by B instead of A. We see that the true effectiveness is much stronger than reported from the RCT. This is because of the lack of external validity in the RCT approach.

Lederer and colleagues draw an important distinction between the goals of prediction and causal inference. We humans are born to be causal thinkers, intuitively giving what we observe causal interpretations. That said, we try to avoid using causal words when interpreting results from a regression model. Often, we are baffled by paradoxical research findings when certain variables are left in or out of a model equation. Our miseries could be alleviated by understanding the distinction between prediction and causal inference from the standpoint of statistical intuition. If we undertake an observational study to identify a causal relationship, then we must completely account for all confounding factors. Failure to do so will leave the residual error from our regression correlated with the exposure variables in the model. This violates the fundamental assumptions of independent residuals in a regression model, which are critical to ensuring unbiased parameter estimates. The regression model is a powerful tool devised for prediction and will do its best to minimise prediction error. This method produces an accurate prediction model by contributing the effect in the uncontrolled confounding to its correlated exposure variables and thus the model coefficients. While it is typically impossible to account for all present confounding, in an RCT setting, the randomisation device serves to remove the confounding effect (both measured and unmeasured); therefore, the prediction model could be used for

causal purposes. This is the only instance in which prediction and causation are equivalent.

Can we understand a causal question using observational data? Answers to this question are still under debate in many contexts, although seminal demonstration has occurred in observational studies of cigarette smoking and lung cancer.⁴ Despite fast adoption of causal inference methods, such as the theory of Neyman-Rubin's potential outcomes framework,⁵ many questions remain to be clarified. In order to control for higher dimensions of confounding variables, propensity scores have almost become the default approach. Biostatisticians, however, are still debating on how well the propensity score method performs when subjected to model misspecification or when important confounders or higher order terms are left out of models.⁶⁻⁸ Can we really call the results from applying causal inference analyses 'causal'? The guidelines article suggests one may wish to call the results 'casual associations', while others encourage the use of the term 'causal' when necessary methodological rigour is applied to analysing observational data.⁹ The field of causal inference is fast evolving in statistics and epidemiology. Undeniably, the advances in methodologies and publication of guidelines are bringing about further clarity and producing causal interpretation to the findings using observational data, similar, in some respects, to a well-conducted RCT study. As we continue sharpening causal inference methods and advance our ability to collect well-measured variables that allow us to minimise the impact of unmeasured confounders in observational settings, we may soon find that we can confidently regard the statistical results from an observational study as 'causal'.

Correction notice This article has been corrected since it was published. Major changes have been made to the text in the body of the article and Table 1.

Acknowledgements The authors thank Dr David Lederer for his discussion with them on the guidelines and specifically the term 'causal association'.

Contributors RS and BH conceived of the presented ideas and commentary. Both authors contributed to the content and approved the final manuscript.

Funding This work was supported in part by grants ME1408-19894 from the Patient Centered Outcomes Research Institute (BH), SZCES18Y7 from the Cystic Fibrosis Foundation (BH) and NIH/NHLBI grant K25 HL125954 (RS).

Competing interests RS serves as a statistical editor on the *Thorax* Editorial Board.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

Confounding Model

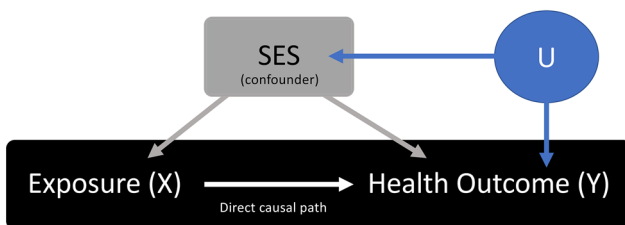


Figure 2 Confounder model as shown using a directed acyclic graph. SES, socioeconomic status.



To cite Huang B, Szczesniak R. *Thorax* 2019;**74**:323–325.

Accepted 14 January 2019



► <http://dx.doi.org/10.1136/thoraxjnl-2018-212488>

Thorax 2019;**74**:323–325.
doi:10.1136/thoraxjnl-2018-212780

REFERENCES

- 1 Lederer DJ, Bell SC, Branson RD, *et al.* Control of confounding and reporting of results in causal inference studies. Guidance for authors from editors of respiratory, sleep, and critical care journals. *Ann Am Thorac Soc* 2019;**16**:22–8.
- 2 Quint JK, Minelli C. Can't see the wood for the trees—confounders, colliders and causal inference: a clinician's approach. *Thorax* 2019;**74**:321–2.
- 3 Wagner CH. Simpson's paradox in real life. *The American Statistician*. 1982;**35**:46–8.
- 4 Cornfield J, Haenszel W, Hammond EC, *et al.* Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst* 1959;**22**:173–203.
- 5 Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc* 2011;**100**:322–31.
- 6 Tsiatis AA, Davidian M. Comment: demystifying double Robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci* 2007;**22**:569–73.
- 7 Gutman R, Rubin DB. Estimation of causal effects of binary treatments in unconfounded studies. *Stat Med* 2015;**34**:3381–98.
- 8 King G, Nielsen R. *Why propensity scores should not be used for matching*. 2016.
- 9 Hernán MA. The C-Word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health* 2018;**108**:616–9.

Correction: *Can't see the wood for the trees: confounders, colliders and causal inference- a statistician's approach*

Huang B, Szczesniak R. Can't see the wood for the trees: confounders, colliders and causal inference- a statistician's approach. *Thorax* 2019;74:323–325. doi: 10.1136/thoraxjnl-2018-212780

Significant changes have been made to the text of this article, including Table 1, since it was first published online and in print. The correct version is available online here: <http://dx.doi.org/10.1136/thoraxjnl-2018-212780>

© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

Thorax 2019;74:e2. doi:10.1136/thoraxjnl-2018-212780corr1

