1

Collider Bias and Mediation Analysis

The main text of Chapter 2 introduced the concept of a collider variable and discussed key properties of it. This document provides a more in depth analysis of colliders in mediation models and how to deal with them. I use DAGs instead of traditional influence diagrams to illustrate key points because DAGs, in my opinion, strike a more compelling visual for the analogies I invoke. They also are more parsimonious because they omit disturbance terms, which are implied rather than explicit. Finally, some of the terminology in DAGs make it easier to explain concepts for colliders.

I use as my primary example a classic mediation model posed by Jacob Yerushalmy (1971), which is shown in Figure 1. The model examines the effects of women smoking during pregnancy on the birthweight of their babies which, in turn, impacts infant mortality. Birthweight is a partial mediator of the effect of smoking on mortality. Path *a* reflects the effect of smoking (X) on birthweight (M), path *b* is the effect of birthweight on child mortality (Y), and path *c* is the direct effect of smoking on child mortality independent of its effect on birthweight.

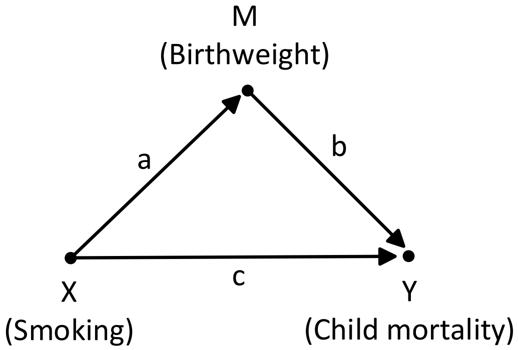
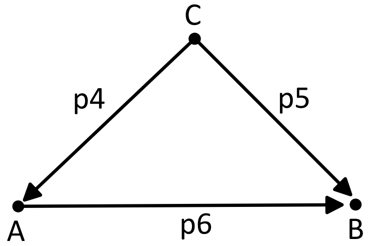


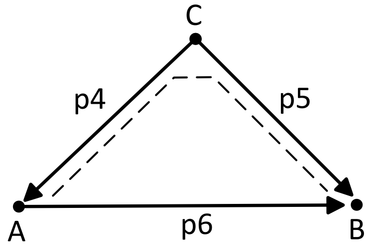
FIGURE 1. Mediation model of smoking and infant mortality

When evaluating this model, Yerushalmy found what he called a "birth-weight paradox," namely that the direct effect of smoking on mortality holding birthweight constant (path *c*) was negative, implying that smoking during pregnancy had a beneficial effect on infant mortality. This paradox went unexplained until the early 2000s (Roth, 2023).

To understand the operative dynamics, I briefly review the way in which we think of confounds and causal inferences using DAGs. Consider the case where I want to infer the strength of the causal relationship, p6, between two variables, A and B, in the presence of a confound, C, per the following diagram:

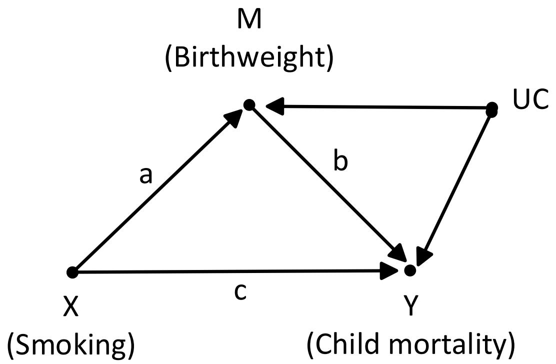


A general strategy in causal modeling is to infer the strength of a causal effect between two variables based on the magnitude of the association between the variables *absent the effects of any confounds* on them. Pearl and Mackenzie (2018) discuss causal and non-causal associations between variables using an analogy that conceptualizes the lines connecting two variables in a DAG as pipes through which causal and non-causal information flows. Causal effects "flow" through the pipes in the directions that the causal arrows point. However, non-causal associational information flows through the pipes in both directions, creating a connection between the two variables that are linked but a connection that is non-causal. I augment the above diagram with a dashed line to graphically portray this latter dynamic (although this heuristic device is not typically used with DAGs).:

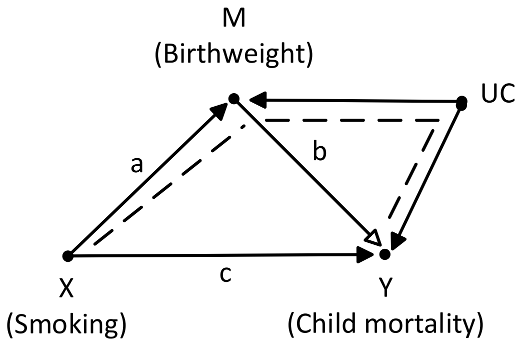


In this figure, the correlation between A and B has two sources, (a) the causal flow of A to B via p6 and (b) the non-causal association flow connecting A and B via p4 and p5 (the dashed line). Pearl and Mackenzie (2018) refer to the causal flow through p6 as going into B from A through the open "front door" of B. The non-causal associational flow through p4 and p5 is said to come into B from A through the open "back door" of B. The idea when modeling data is to cut off or "block" all relevant back door flows that link two variables; the causal flow through the front door will then reflect the association that remains. In a statistical analysis, for example, I might block the flow of non-causal associational information through a given back door by controlling for (or holding constant, or covarying out) the variable through which the back door flow occurs, in this case variable C. This would be accomplished by including C as a covariate when regressing B onto A.

Returning to the Yerushalmy model in Figure 1, suppose there is an unmeasured confound for the M→Y link, as follows (where UC stands for an unmeasured confound):



Note that the presence of this confound turns the birthweight mediator, M, into a collider (because M is influenced by both X and UC). If we hold the collider, M, constant when estimating path *c,* which is a common strategy to estimate the direct effect of X on Y or path *c*, it turns out that we unwittingly open a back door non-causal association between X and Y through UC (see Pearl, 2009, for a mathematical proof). This, in turn, introduces bias into the estimate of path *c*. Here is the dynamic shown graphically in the DAG:



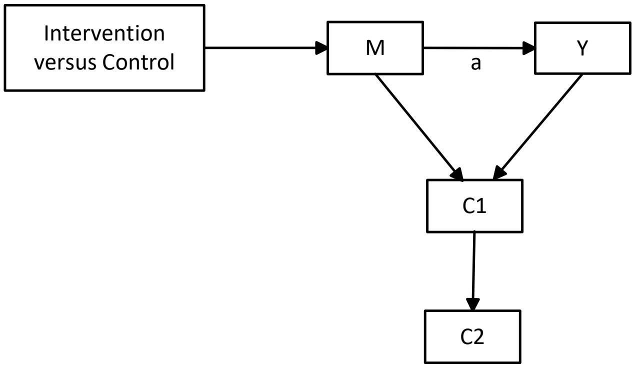
A solution to the problem created by the presence of UC is to identify the UC variable when planning your RET, to measure it during data collection (I will use the acronym MC for the measured version of UC), and then hold MC constant when predicting Y to close the back door that covarying out the collider M opened. That is, close the back door per the equation that predicts Y from X, M and MC. In this latter equation, the coefficient for X→Y will now be an unbiased estimate of the direct effect of interest unencumbered by the collider bias. The birthweight paradox of Yerushalmy would disappear if we can identify UC, measure it, and control for it.

I noted in the main text that some scientists recommend against controlling for post-treatment variables because doing so can introduce collider bias. This recommendation undermines facets of mediation analysis because the mediators are, after all, post-treatment variables. Note that if I can reasonably assume (a) there are no unmeasured confounds (UCs) operating, or (b) that the impact of the UCs are weak enough that they are not consequential, or (c) that there are multiple UCs whose effects cancel or nearly cancel each other, then the problem of colliders interfering with mediation analysis disappears.[1] Also, when planning my RET, if I can identify moderate to strong UCs so I can measure and control them to remove bias, then the problem of colliders also disappears. As discussed in the main text of Chapter 2, perhaps I cannot identify and control all of the relevant UCs, but if I am able to do so for the major ones, those that I omit may be inconsequential.

The bottom line is that like most statistical modeling, mediation analysis carries assumptions with it. The key assumption in this case is that there are no *unmeasured* confounds (UCs) that create non-trivial coefficient bias via collider dynamics or otherwise. The OLS regression counterpart to this assumption is the well-known assumption of no omitted variables (see Chapter 5). When you conduct standard regression modeling, you bring with it a set of assumptions that are necessary to make inferences.

A final point worth noting is the recognition that if you do not hold constant a collider, you can still obtain collider bias if you statistically control for a descendant of the collider (recall from the main text of Chapter 2 that a descendant of a target variable is a variable that is impacted by that target variable “downstream”). Here is an example of a model with a collider descendant using a traditional influence diagram (omitting disturbance terms):

[1] Greenland (2003) notes that collider bias often will be weak unless the causal links creating the collision are quite strong. For a discussion of strategies for estimating the magnitude of collider bias, see Greenland (2003), Groenwold, Palmer and Tilling (2021), Nguyen, Dafoe and Ogburn (2019), and Whitcomb, Schisterman, Perkins, and Platt (2009).



The variable C2 is a descendant of the collider C1. If I hold C2 constant when I estimate the effect of M on Y (but omit C1), I will still introduce bias by creating a backdoor link between M and C2.

In RETs, the potential for collider bias is likely to arise when there are multiple mediators with causal relationships among those mediators or multiple outcomes with causal relationships among those outcomes (see Novak, Boutwell & Smith, 2024).

REFERENCES

Golding, J. (2024). A brief introduction to colliders. Institute News https://jeangoldinginstitute.blogs.bristol.ac.uk/2019/10/28/a-brief-introduction-to-colliders/

Greenland, S. (2003). Quantifying biases in causal models: Classical confounding vs. collider-stratification bias. *Epidemiology*, 14, 300-306.

Groenwold, R., Palmer, T. & Tilling, K. (2021). To adjust or not to adjust? When a "confounder" is only measured after Exposure. *Epidemiology*, 32, 194–201.

Nguyen, T., Dafoe, A. & Ogburn, E. (2019). The magnitude and direction of collider bias for binary variables" *Epidemiologic Methods*, 8, 20170013.

Novak, A., Boutwell, B. & Smith, T. (2024). Taking the problem of colliders seriously in the study of crime: A research note. *Journal of Experimental Criminology, 20*, 1005–1014.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Roth, L. (2023). Pathway analysis, causal mediation, and the identification of causal mechanisms. In A. Damonte and F. Negri (Eds.) Causality in policy studies. (pp123-158). Springer.

Whitcomb, B., Schisterman, E., Perkins, N. & Platt, R. (2009). Quantification of collider-stratification bias and the birthweight paradox. *Paediatric and Perinatal Epidemiology*, 23, 394–402.

Yerushalmy, J. (1971). The relationship of parents' cigarette smoking to outcome of pregnancy – Implications as to the problem of inferring causation from observed associations. *American Journal of Epidemiology, 93*, 443–456.