

Appendix S1

“Applying the structural causal model (SCM) framework for observational causal inference in ecology” *Ecological Monographs*

Suchinta Arif and Aaron MacNeil

Preventing Overcontrol, Confounding and Collider Bias:

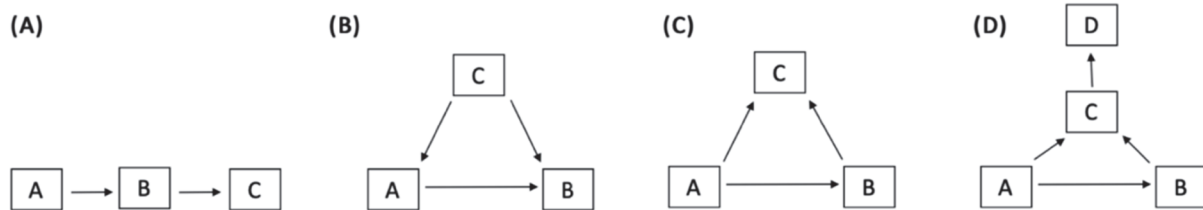


Figure S1. DAGs depicting a (a) chain, (b) fork, (c) collider, and (d) descendant of a collider. In order to estimate the effect of A on B, biases associated with these structures, which include confounding, overcontrol, and collider bias, must be avoided.

The components of a DAG can be broken down into three types of causal structures: chains, $A \rightarrow C \rightarrow B$ (Fig. S1a); forks, $A \leftarrow C \rightarrow B$ (Fig. S1b); and colliders $A \rightarrow C \leftarrow B$ (Fig. S1c). When appropriate statistical adjustments are not made, these three causal structures can each lead to a specific type of bias: overcontrol, confounding, and collider bias (e.g. see also Elwert, 2014). Here, we review how the backdoor criterion directs us to avoid these biases, which can otherwise plague observational correlative studies.

Overcontrol bias: In a chain, $A \rightarrow C \rightarrow B$, two variables may be associated because one variable, A, indirectly causes the other, B (Fig S1a). If we want to find the effect of A on B, conditioning on C would block the association flowing from A to B. This is known as ‘overcontrol bias’ and can be resolved by not conditioning on an intermediate variable between predictor and response variable. Here, the application of the backdoor criterion instructs us to not condition on variable C in order to determine the effect of A on B.

Confounding bias: In a fork, $A \leftarrow C \rightarrow B$, two variables, A and B, may be associated because of a common cause, C (Fig S1b). If we want to find the effect of A on B, then not conditioning on C would cause a spurious, or biased association between A and B. This is known as ‘confounding bias’ and can be resolved by conditioning on the common cause. Here, the application of the backdoor criterion instructs us to condition on variable C in order to determine the effect of A on B.

Collider bias: In a collider, $A \rightarrow C \leftarrow B$, two variables, A and B, may be associated because they have a common outcome, C (i.e. a collider, Fig S1c). Conditioning on a collider variable, C, creates a spurious, or biased association between A and B. This is known as ‘collider bias’ and can be resolved by not conditioning on a collider (e.g. C in Fig S1c and S1d) or any descendant of a collider (e.g. D in Fig S1d). Here, the application of the backdoor criterion instructs us not to condition on variable C (Fig S1c and S1d) or D (Fig S1d) in order to determine the effect of A on B.

While examples of confounding variables can be found throughout the ecology literature (*e.g.* land use change acts as a confound for determining the effect of climate on elevational species redistribution, Guo et al., 2018), the same cannot be said for overcontrol and collider bias. A literature search using Web of Science with the search terms “ecology” and “overcontrol” or “collider” resulted in no papers that mentioned these biases. This does not mean that these biases do not occur in observational ecological studies, but rather that they are not knowingly adjusted for. It is also noted that while ecologists are aware of confounding, it does not mean they are

accurately being adjusted for across observational studies. For example, covariate adjustments resulting in the inclusion of multiple potential confounders can lead to bias, instead of reducing it (Shrier and Platt, 2008).

Given a DAG, the application of the backdoor criterion will eliminate overcontrol, confounding, and collider bias, allowing for more reliable causal estimates from observational data.

Literature Cited:

Elwert, F. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, 40, 31-53. Doi: 10.1146/annurev-soc-071913-043455

Guo, F., Lenoir, J., Bonebrake, T. (2018). Land-use change interacts with climate to determine elevation species redistribution. *Nature Communications*, 9, 1315. doi: 10.1038/s41467-018-03786-9

Shrier, I., Platt, R. (2008). Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology*, 8, 70. doi:10.1186/1471-2288-8-70