Collider Bias

Mathias J. Holmberg, MD, MPH, PhD; Lars W. Andersen, MD, MPH, PhD, DMSc

Bias is a systematic, nonrandom error in the estimation of a treatment effect or the effect of an exposure or risk factor. Bias can lead to invalid results in observational studies and randomized clinical trials (RCTs). Bias is often broadly categorized into 3 groups: confounding, information (or measurement) bias, and selection bias. Selection bias is a general term describing bias that occurs when study participants are identified in a manner such that they are no longer representative of the target population. This can occur when an exposure and outcome each influence a common third variable—the collider—and that variable has been controlled for in the statistical analysis of the study data. Collider bias threatens the internal validity of a study and the accurate estimation of causal relationships.

In an observational study of 4480 patients with confirmed COVID-19 published in JAMA, Fosbøl et al found no increase in mortality among patients using angiotensin-converting enzyme inhibitors (ACEIs) or angiotensin receptor blockers (ARBs). The possibility of collider bias should be considered in interpreting this result because the study was restricted to patients with COVID-19, and COVID-19 might represent a collider associated with drug treatment and mortality.

What Is Collider Bias?
Collider bias occurs when an exposure and outcome (or factors causing these) each influence a common third variable and that variable or collider is controlled for by design or analysis. In contrast, confounding occurs when an exposure and outcome have a shared common cause that is not controlled for. Methods for statistically controlling for a variable include restricting the analysis to patients with a given characteristic (ie, the patients have been selected for this analysis) or applying a statistical adjustment based on a variable (eg, the variable of interest is included as a variable in a regression model). Collider bias is often inadvertently introduced by controlling for a variable that occurs after the exposure or intervention.

Collider bias can be illustrated using directed acyclic graphs (DAGs). A DAG is a graphical representation of the potential causal relationships between variables, with arrows used to denote the direction of causality. Collider bias occurs when 2 arrows collide on a variable that has been controlled for (panel A in the Figure).

This collision creates a spurious or artificial association between the 2 other variables (ie, A and B). Consider an extreme example in which the outcome (labeled as C in panel A) can only be caused by either of the 2 unrelated variables (ie, A or B) but not the combination of A and B. If the analysis is restricted to patients who had the outcome of C, patients either have A or B, but not both, and a spurious (negative) association is created between A and B. This might occur, for example, in an RCT that requires each patient to have 1 of 2 risk factors, but not both, because of the intent to enroll a medium-risk population (eg, because there already is a therapy for higher-risk patients with both risk factors). Then, if the enrolled population was used to examine the relationship between the 2 risk factors, a spurious (negative) association between them would be found.

Why Is Avoidance of Collider Bias Important?
Collider bias is a threat to validity in observational studies and RCTs. It is often less readily recognized than confounding.

A study by Valls-Pedret et al illustrates an example of potential collider bias in an RCT. In this secondary analysis of data from an RCT that compared 2 Mediterranean diets and a control diet in 447 participants at high cardiovascular risk, cognitive function was improved with both dietary interventions. However, loss to follow-up was higher in the control group (33%) as compared with the Mediterranean diet groups (16% and 23%). Differential loss to follow-up can introduce collider bias because the analysis is restricted to patients who are not lost to follow-up (ie, only those with follow-up data are included in the analysis) (panel B in the Figure). If participants assigned to one of the Mediterranean diets were less likely to be lost to follow-up than those in the control group (which is supported by the trial) (arrow 1 in panel B) and participants with poor cognitive function were independently more likely to be lost to follow-up (arrow 2 in panel B), restricting the analysis to those not lost to follow-up could result in a spurious or noncausal association between diet and cognitive function. In contrast to what the trial
found, these associations would result in bias toward finding a benefit in the control group. As such, collider bias does not explain the results of the trial but illustrates how collider bias can occur in an RCT (eg, through loss to follow-up).

**Alternatives to Study Designs Prone to Collider Bias**

Awareness of the potential for collider bias and avoiding study designs or statistical analyses that are prone to collider bias are important in observational studies and RCTs. Addressing collider bias is best done during the design of a study, for example by minimizing loss to follow-up or avoiding restricting the study population based on characteristics likely to be affected by both the exposure and outcome of interest. DAGs may be helpful in exploring potential causality and identifying collider bias. It is important to identify and distinguish collider bias from other types of bias and consider how the choice of study design and statistical analysis may introduce, increase, or reduce bias. For example, the choice of statistical method can account for confounding, while at the same time introduce collider bias.

**How Does Consideration of Collider Bias Apply to the Study by Fosbøl et al?**

In their study of patients with COVID-19, Fosbøl et al found no increased mortality among patients using ACEIs or ARBs. However, it has been hypothesized that ACEI or ARB use could result in increased susceptibility to COVID-19. This is true, collider bias is a concern in this study because the analysis was restricted to patients with confirmed COVID-19 (panel C in the Figure). If ACEI or ARB use increases the risk of COVID-19 (arrow 1 in panel C) and other unrelated risk factors cause COVID-19 (arrow 2 in panel C), restricting the analysis to patients with COVID-19 will create a spurious negative association between ACEI or ARB use and these risk factors. If these risk factors are also related to mortality (arrow 3 in panel C), a spurious association between ACEI or ARB use and mortality could appear. In other words, by requiring the participants to have COVID-19 for inclusion in the analysis (ie, controlling on that characteristic), a spurious negative association could be generated between the use of ACEIs or ARBs and risk factors for COVID-19. Since these same risk factors are associated with mortality, this, in turn, creates a spurious protective association between ACEI or ARB use and mortality.

**How Does Consideration of Collider Bias Influence Interpretation of the Study by Fosbøl et al?**

Fosbøl et al addressed the potential for collider bias arising from limiting the study to patients with COVID-19 by showing that ACEI or ARB use was not associated with susceptibility to confirmed COVID-19. Thus, arrow 1 in panel C can be removed, and there is no longer a collision of causal relationships affecting COVID-19 illness.