

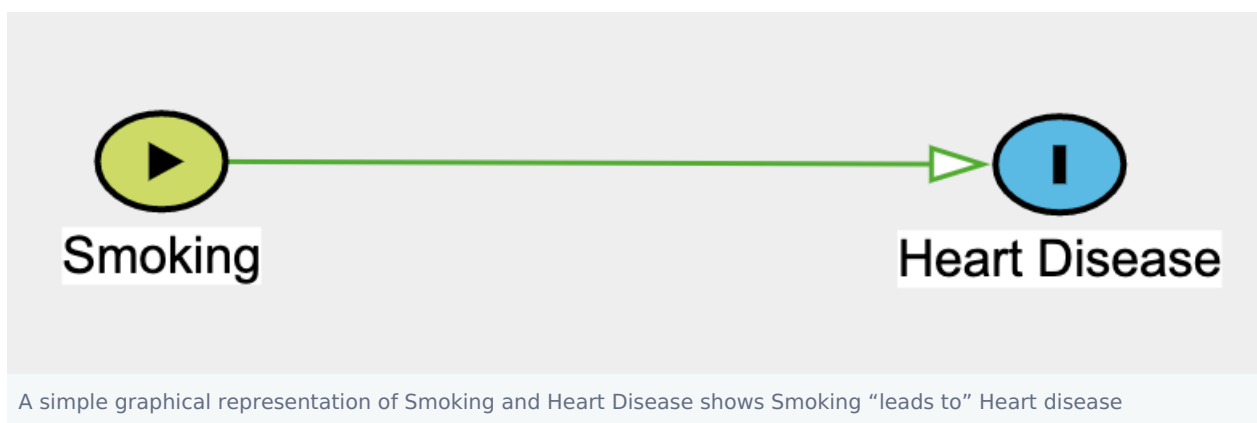
Concept of confounding variables

Arindam Basu¹

¹ University of Canterbury

Potential competing interests: The author(s) declared that no potential competing interests exist.

Confounding variables in Epidemiology indicate the phenomenon that happens due to the presence of a “third variable” or a “third set of variables” in the association between an Exposure and an Outcome that “confuses” any true association between the Exposure and Outcome. As an example, imagine a researcher were to study whether smoking causes heart disease, and tested the theory that compared with those who smoke two or fewer packs of cigarette per day, those who smoke three or more packs of cigarette per day, are more likely to suffer from heart disease (heart disease that result from reduced flow of blood to the arteries that feed heart muscles). Using graphical terms, we represent this as follows:



This above diagram shows a graphical presentation of Smoking “causes” heart disease or smoking “leads to” heart disease. Hence the arrow is headed FROM Smoking TO Heart Disease. This also means that measurements of smoking and measurements of heart disease will be associated with each other. But the question is, is this association “REAL”?

Now, add to this the fact that we know heart disease follows a different pattern between men and women ^[1], so sex will have an impact on heart disease independent of smoking, and we represent this as follows:

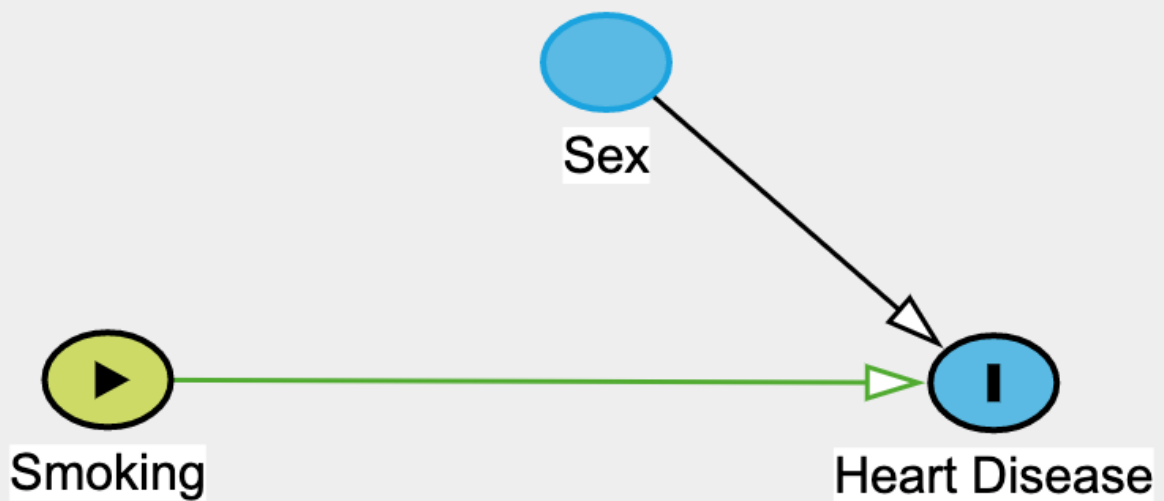
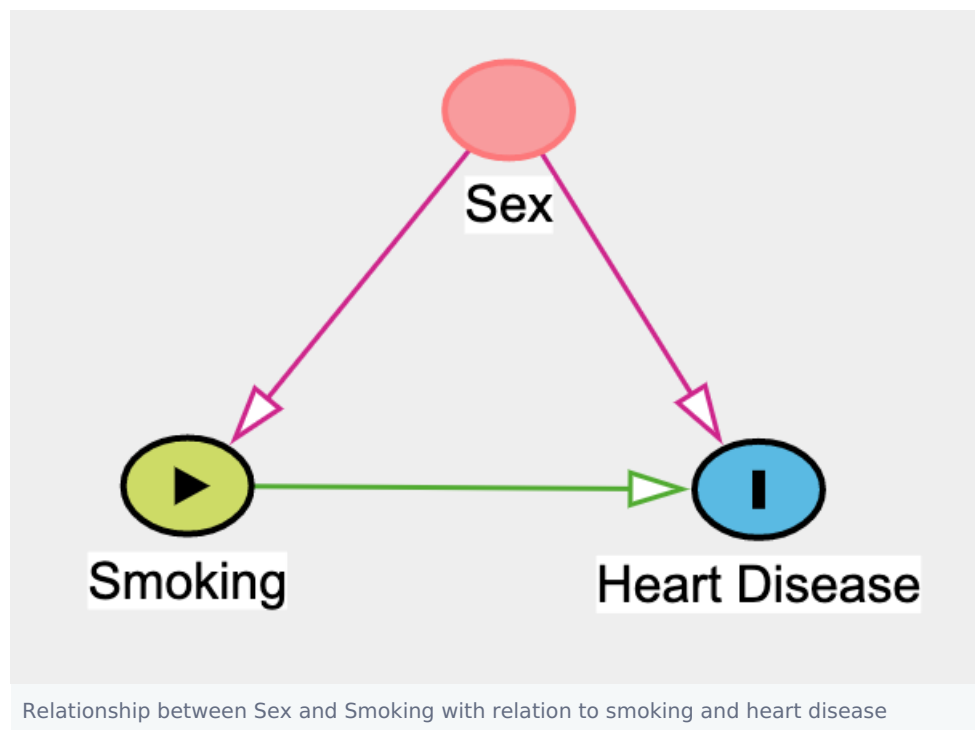


Illustration that Sex has an impact on heart disease independent of smoking

But we also know that the smoking patterns of men and women are different^[2], so now we add another arrow that goes FROM Sex TO Smoking to indicate that Sex differences in smoking patterns. So, now we have this:



Relationship between Sex and Smoking with relation to smoking and heart disease

So, now we have three arrows connecting:

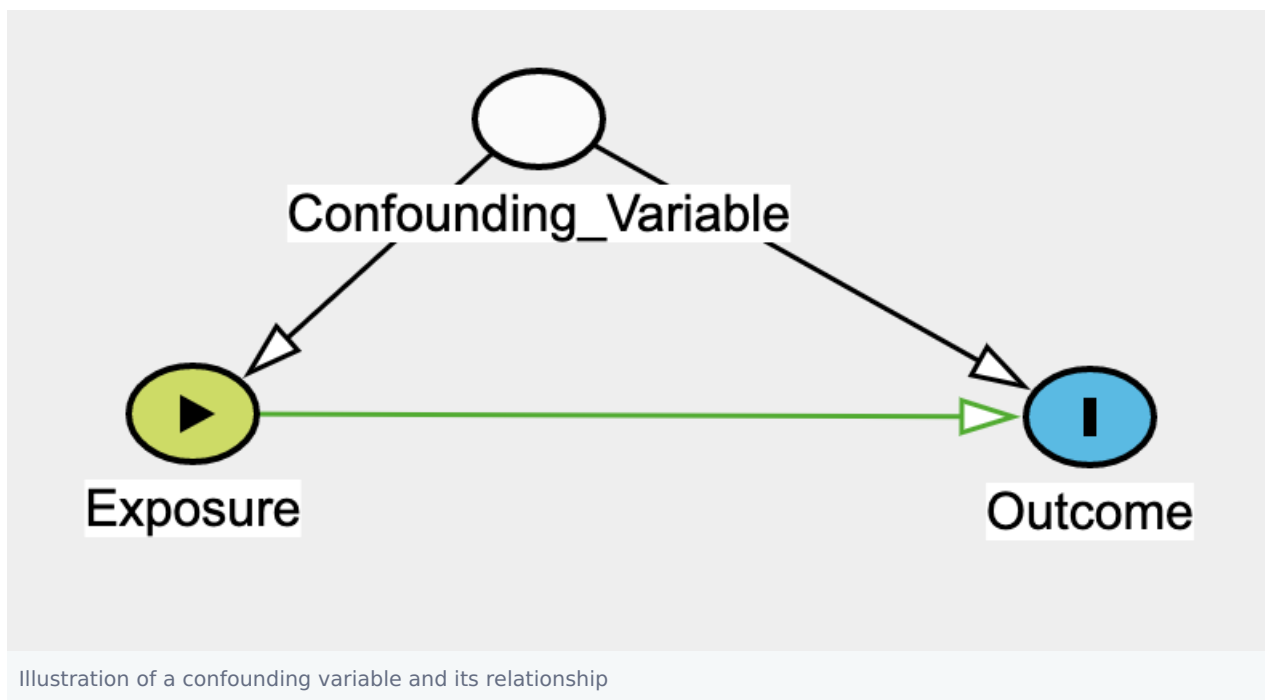
1. An arrow that indicates that Smoking is a possible cause of heart disease,
2. An arrow that indicates that Sex is associated with variations in heart disease

3. An arrow that indicates that Sex is also associated with variations in the patterns of smoking

Note that when we define these relationships, we assume that:

1. Smoking does NOT lead to Sex (this is an absurd idea) that then leads to heart disease
2. We could as well state that the relationship between smoking and heart disease that we get to study here might be actually attributed to Sex rather than Smoking, so Smoking is NOT INDEPENDENT of Sex when we explain heart disease risk.

This is the reason, in this case, we state that Sex is CONFOUNDING any observed association between Smoking and Heart Disease. So, now we state a more general case and show it thus:



So, a confounding variable is one which is associated with both the Exposure (or Intervention) and Outcome, but does NOT come in any causal pathway that can connect the exposure (or Intervention) and the outcome. The point to note here is that, Confounding is decided on “conceptual” basis, and that, there can be more than one confounding variables or sets of confounding variables. It is possible to create more complex pathways that connect the exposure and outcome via mechanisms of confounders.

What need to be done about confounding variables?

Confounding variables are decided on the basis of available knowledge and treated as such. There are generally three ways in which confounding variables are addressed in health research projects.

- **Before commencing a study**, if there are known confounding variables, and if the study is an observational epidemiological study (such as case control study or cohort study), those with the exposure (i.e., exposed or non-exposed cohorts) OR with the disease (i.e., cases and controls) are MATCHED. For example, if you were to conduct a study on the health risks of exposure to inorganic arsenic and lung disease, and you considered AGE as a confounding variable, then you would take roughly same number of people who have lung disease and who do not have lung disease. This would

be “matching” on age variable.

- If you are conducting an intervention study, then you conduct **randomisation** so that those who will receive the intervention and those who will receive the control condition or placebo, or alternative intervention, are randomly allocated their interventions. In this situation, you first use a random numbers table AND THEN allocate the participants to either the intervention or the control arm. The advantage of this procedure is you can control for both known and unknown confounding variables.
- Once you have collected your data, you can control for confounding by conducting **stratified analysis** of the data, or using **multivariable analysis** of the data. These are statistical methods of adjusting for the presence of confounding variables.

References

1. ^Lori Mosca, Elizabeth Barrett-Connor, Nanette Kass Wenger. (2011). *Sex/Gender Differences in Cardiovascular Disease Prevention*. *Circulation*, vol. 124 (19), 2145-2154. doi:10.1161/circulationaha.110.968792.
2. ^Stephen T. Higgins, Allison N. Kurti, Ryan Redner, Thomas J. White, et al. (2015). *A literature review on prevalence of gender differences and intersections with other vulnerabilities to tobacco use in the United States, 2004–2014*. *Preventive Medicine*, vol. 80 , 89-100. doi:10.1016/j.ypmed.2015.06.009.