Causal Diagrams

Taylor Okonek and Charlie Wolock

BIOST 311, Spring 2022

**Introduction & Terminology**

Often, when thinking about the need to adjust for additional covariates, it is helpful to draw a causal diagram relating the outcome, predictor of interest, and additional covariates. Diagrams consist of the following parts:

* Nodes: variables (including the outcome, predictor of interest, and additional covariates)
* Edges: connections between nodes, to denote causal relationships or associations
  + Line: denotes that two variables are associated with one another
  + Arrow: denotes that one variable is *causally related to* another

**Example**

Suppose we observe Seattleites every day for a year, and observe a positive relationship between the number of umbrellas used on a given day and the number of car accidents. In particular, we note that the more umbrellas used, the more car accidents there are on a given day. Taylor decides that this must be a causal relationship, and so she makes the following diagram for this relationship:

Number of car accidents

Number of umbrellas used

Note that the arrow indicates that Taylor believes umbrellas are causally related to car accidents (in this case, more umbrellas used leads to more car accidents).

Charlie thinks there’s an important variable at play here that Taylor is ignoring: rain. He thinks that the presence of rain on a given day (1) increases the number of umbrellas used and (2) increases the number of car accidents. Charlie makes the following diagram for this relationship:

Number of car accidents

Number of umbrellas used

Rain

He thinks that the relationship between number of umbrellas used and number of car accidents is *not* causal, but instead just an association. He does, however, think that there are causal relationships between rain and number of umbrellas used, and rain and number of car accidents, as indicated by arrows instead of lines.

**Causal pathways**

A causal pathway indicates a path between nodes in a causal diagram, directed using arrows. *A pathway cannot be causal if there are no arrows*. In the previous example, there were (short) paths between rain and number of car accidents, and rain and number of umbrellas used. Causal diagrams can also be much more complicated, with much longer paths:

H

B

C

A

I

F

E

D

G

Suppose we’re interested in finding causal pathways from node A to node G in the above graph. One such pathway is A -> D -> G. Note that the relationship between B and C is not causal since there is no arrow, and so A -> B - C -> F -> G is not a causal pathway from A to G. Note also that the relationship from A <- I -> G is not a causal pathway because the arrows are not all pointing in the same direction, ending with G.

1. List the other causal pathways from A to G.
2. List the causal pathways from I to G.

**Causal pathways practice**

1. Suppose Taylor instead believed that car accidents caused people to use more umbrellas (perhaps there is some superstition that keeping an umbrella in your car will prevent an accident from happening). Fill in the casual diagram for this relationship below.

Number of car accidents

Number of umbrellas used

1. Suppose scientists are interested in the relationship between anxiety and jaw tension in adults. In particular, they hypothesize that increased anxiety leads to increased jaw tension (a causal relationship). Other variables they consider to be relevant include:

* Grinding teeth at night
* Wearing a night guard
* Wearing headgear as a child
* Playing contact sports as a child
* Daily meditation

Fill in the causal diagram below with any relationships between variables that you think may be present.

?

Jaw tension

Anxiety

Night guard

Grind teeth

Meditation

Contact sports

Headgear

1. For the causal diagram above, explain any lines or arrows you may have drawn between nodes. Why do you think certain relationships may exist between variables?
2. Suppose scientists are interested in the relationship between access to antenatal clinics and maternal mortality. Antenatal clinics provide care from health care individuals during pregnancy. They hypothesize that pregnant individuals in areas with *easy* access to antenatal clinics will have lower mortality rates, assuming that individuals in fact use the services provided by the clinics. They know of at least a few variables associated with antenatal clinic availability (urban/rural location, distance to nearest clinic, ability of the patient to pay for care, other type of pregnancy care more easily available), and a few variables associated with increased maternal mortality (delivery method, age, state of anemia, pregnancy complications), but the scientists are having trouble determining how everything relates to each other, and are worried they might be missing some important variables. Fill in the causal diagram below with any relationships between variables that you think may be present.

Antenatal clinic access

Maternal mortality

?

**Confounders**

The reason why causal diagrams are so nice is that they can help us easily identify confounding variables, effect modifiers, and precision variables. We’ll touch on confounding first. Recall:

* Confounder (or, “confounding variable”): a variable that is *causally related to* our outcome and also *associated with* the exposure in our sample

In a causal diagram, a confounder looks like this:

Predictor of Interest

Confounder

Outcome

We can see that the confounder both causes the outcome (denoted with an arrow) and is associated with the predictor of interest (denoted by a line). A variable can also be a confounder if it causes *both* the outcome and predictor of interest, like this:

Confounder

Predictor of Interest

Outcome

HOWEVER, the direction of the arrow matters here. If the predictor of interest *causes* the potential confounder, then it is in fact *not* a confounder.

Predictor of Interest

Outcome

NOT a Confounder

In this case, the variable is *on the causal pathway* between the predictor of interest and the outcome. A confounder cannot be on the causal pathway from the predictor of interest to the outcome.

Recall that we don’t need to worry about confounding in randomized trials, because individuals are randomly assigned to treatment or control. Because individuals are randomly assigned to treatment or control, the treatment variable *cannot* be associated with any other variable (unless randomization fails). This is why we can make causal statements with randomized trials (and hence why they are so useful!)

Outcome

Treatment

not possible!

Any other variable

**Confounders practice**

1. In the example on page 1, Charlie hypothesized that the following relationship existed:

Rain

Number of car accidents

Number of umbrellas used

What is the role of Rain here?

1. Taylor is currently on a mission to get her cat, Alice, to sleep through the night (in an attempt to therefore sleep undisturbed herself). Taylor thinks that if she plays with Alice before bed (predictor of interest), Alice will sleep longer (outcome). Taylor has also noticed in the past that when Alice gets a bedtime snack, she seems to sleep better throughout the night. However, Taylor is worried that feeding Alice a bedtime snack may be a potential confounding variable, since she usually feeds Alice after playing with her so that she can recover her energy (and because Alice is always hungry). Draw a causal diagram for this problem, and determine whether or not Taylor should be concerned that feeding Alice a bedtime snack is a confounding variable.

**Effect Modifiers**

Effect modifiers do exactly what you would expect based on the name: modify the effect of a predictor on the outcome.

* Effect modifier: a variable that modifies the association between the predictor of interest and the outcome. In other words, the association between the predictor of interest and the outcome *depends on* the effect modifier

In a causal diagram, an effect modifier looks like this:

Predictor of Interest

Outcome

Effect modifier

This is the first time we’ve seen an arrow pointing to another edge! The reason we do this for effect modifiers is because they alter the *association* between two variables, not necessarily either variable on its own.

**Effect modifier practice**

1. Scientists at UW have developed a new drug called Helpmesleep, which has been shown in studies of mice to lengthen duration of sleep. They believe the drug may be particularly beneficial for individuals with insomnia in improving quality of life, and are beginning to plan a randomized trial to test average sleep duration in individuals with insomnia given Helpmesleep vs. those given a placebo. In preliminary studies with mice, the scientists noticed that sleep duration was *longer* when the mice were given caffeine in addition to Helpmesleep, but shorter when the mice were given only caffeine. In lab tests of the drug, they noticed that the chemical structure of caffeine served to increase the strength of Helpmesleep. Draw a causal diagram relating the following variables: caffeine, sleep duration, and Helpmesleep.

What role do each of these variables play (options: predictor of interest, outcome, confounder, effect modifier)?

1. Xeroderma pigmentosum (XP) is a genetic condition characterized by extreme sun sensitivity. It is well established that exposure to ultra-violet (UV) radiation from the sun increases the risk of skin cancer. Taylor and Charlie each draw causal diagrams relating the variables XP, UV exposure, and skin cancer risk.

Taylor’s diagram:

Increased skin cancer risk

UV exposure

XP

Charlie’s diagram:

XP

Increased skin cancer risk

UV exposure

Explain the reasoning behind Taylor’s diagram.

Explain the reasoning behind Charlie’s diagram.

Which diagram do you think is correct, and why?

**Precision variables**

The term “precision variable” is something often talked about in the UW statistics and biostatistics departments, but is not commonly used language across statistics as an entire field. Nevertheless, we think it’s a useful distinction, and so we’ll talk about it in this course!

* Precision variable: a variable that is associated with (or causes) the outcome *in the population*, but is not associated with the predictor of interest *in the sample*

In a causal diagram a precision variable looks like one of the following:

Predictor of Interest

Outcome

Precision variable

Predictor of Interest

Outcome

Precision variable

Note that the precision variable is not associated with the predictor of interest in any way, and is only associated with (or causes) the outcome.

Precision variables serve to *reduce the standard error* of our estimates, in particular, our estimate of the relationship between the predictor of interest and the outcome. Smaller standard errors = smaller confidence intervals, which is usually a good thing.

**Precision variable practice**

1. In the example on page 1, Charlie hypothesized that the following relationship existed:

Number of umbrellas used

Number of car accidents

Rain

Redraw the causal diagram below, imagining that rain is a precision variable rather than

a confounder.

What would it mean, in words, for rain to be a precision variable rather than a

confounder? Do you think it makes more sense for rain to be a precision variable or a confounder? Why?

1. Scientists are interested in estimating the effect of having unprotected sex on risk of HIV infection. They hypothesize that having unprotected sex will lead to increased risk infection, but they are specifically interested in *how much* greater the risk of infection is among individuals who have unprotected sex, compared to those who do not. They want their estimate to be as precise as possible, so they consider the following variables as possible candidates for precision variables to include in their model:
   * Needle sharing
   * HIV+ birth parent (parent who gave birth to them was HIV+ while giving birth)
   * Having a STI
   * Having a sexual partner who is at high risk or already has HIV

For each of these four possible candidates, draw what you believe to be an accurate causal diagram. Based on these diagrams, determine whether each of the four candidates is a precision variable or not. Give your scientific reasoning for why you drew any edges in the causal diagrams.