

Establishing Causality

by Sophia



WHAT'S COVERED

This tutorial will explain guidelines for establishing causality. Our discussion breaks down as follows:

1. Causality

2. Levels of Confidence

1. Causality

You might recall that **causality** is a cause-and-effect relationship between variables.

Sometimes you may want to determine whether two variables are well-correlated due to cause and effect. The best way to do it is with a controlled experiment, but sometimes you cannot do a controlled experiment. Perhaps you have to do an observational study due to ethical or practical concerns.

How can you prove cause and effect under those circumstances? It's still possible, though very difficult, to prove cause and effect with a study that isn't an experiment.

The study will need to meet these five criteria:

Criteria	Questions to Ask
Consistency: You need to look for cases when correlation remains while other factors vary.	Does the association remain even when other variables are allowed to vary? Does this work across different races and genders? Do high amounts of the alleged cause lead to high or low amounts of the alleged effect?
Control: You need something similar to a control. It's not exactly using a control group, but it's similar to what you would do if you had	Is the effect absent when the cause is absent? Is the effect present when the cause is present?

done an experiment. This is essentially like splitting a group of volunteers into two groups and having a treatment group and a control group. Although you're not assigning them that way, you're looking for the same thing.	
Correlation: You need to look for evidence that larger amounts of suspected cause produce a larger effect.	Does an increase in the cause correspond to an increase, or hypothetically a decrease, in the effect?
Consideration of Alternatives: You need to check for other possible causes.	Is there be something else, perhaps some lurking variable, that you're missing? Might there be other plausible causes?
Connection: You need to try to determine the physical mechanism for the cause and effect.	What physically might create this effect? What is the physical mechanism behind the effect, and how could it plausibly be led to from the cause?

These are pretty strict requirements. They are necessary in order to determine, without an experiment, whether or not two correlated variables are going to be cause-and-effect related.



THINK ABOUT IT

Consider the following claims and determine if you can establish causality:

Claim: <i>"Eating a lot of carbohydrates makes you gain weight."</i>		
Consistency	✓	Is this consistent across different races, genders, etc? More or less, this claim is consistent.
Control	✗	Is the effect present when the cause is present? Do people who eat lots of carbohydrates gain weight? You can see a lot of people that eat lots of carbohydrates and don't gain a lot of weight.
Correlation	✓	Does an increase in the amount of carbohydrates increase the amount of weight gained?

		All other things being constant--yes, more or less.
Consideration of Alternatives		Is there anything else besides eating lots of carbohydrates that might make people gain weight? It's possible that people that eat lots of carbohydrates don't exercise as much as people that eat fewer carbohydrates. Maybe that's what is making those people gain weight. So since we've considered alternatives and found them to be plausible, we're going to say that we can't say that this is the only cause.
Connection		Is eating lots of carbohydrates physically related to weight gain? They are.

So, this claim almost passed, but it did not meet all of the criteria. So you can't say that this claim is 100% true.

Claim: <i>"Smoking causes lung cancer."</i>		
Consistency		Do you see higher lung cancer rates among smokers across different genders and races? Yes, even across different countries. This is true worldwide.
Control		Is the effect present when the cause is present? Do people who smoke tend to get lung cancer? People can get lung cancer even if they don't smoke. But you see it in much higher rates with people that do smoke, and much lower rates in people that don't smoke.
Correlation		Do groups of people who smoke have higher incidences of lung cancer than people that smoke less?

		Yes, they do.
Consideration of Alternatives	✓	What else might be causing lung cancer? It's possible that there's a genetic link that both causes people to smoke and predisposes them to lung cancer. Although that is somewhat plausible, it isn't highly plausible. Considering the alternatives, you can say that smoking is a more likely cause than genetics.
Connection	✓	Is there a scientifically understood physical connection between smoking and lung cancer? Yes. There have been experiments using the tar in cigarettes on animals, and those animals have developed cancerous tumors. So we understand the physical connection.

This passes all of the criteria so we can reasonably claim that smoking does cause lung cancer. Now, smoking is not going to cause lung cancer in 100% of people. Not everyone who smokes is going to get lung cancer. But we can say this is a large contributor.



TERM TO KNOW

Causality

A cause-and-effect relationship between two variables.

2. Levels of Confidence

You can have different levels in your confidence in the causation. You can have:

- Possible cause, which means you can imagine a scenario where A causes B. One thing causes the other.
- Probable cause, which means you're pretty sure that A causes B.
- Cause beyond a reasonable doubt, which means that you cannot think of a scenario where the response of second variable B could have been caused by anything other than A.

IN CONTEXT

Consider the criminal justice system.

Possible cause would be the case where someone becomes a suspect. There may be evidence to suggest that this suspect committed some crime.

Probable cause would be the instance where the person actually gets arrested for the crime.

Cause beyond a reasonable doubt would be the part where the person is convicted in a court of law.



SUMMARY

The only way to prove 100% definitively causation is with a controlled, randomized experiment. However, by using a set of very stringent criteria, you can reasonably conclude that there's a causal link between two variables based on whether or not they meet five criteria. Sometimes the alleged causes don't hold up under the scrutiny, but we can be certain of the ones that do. For this reason, we can describe levels of confidence in our causation.

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR [TERMS OF USE](#).



TERMS TO KNOW

Causality

A cause-and-effect relationship between two variables.