This tutorial has to do with establishing causality.

Sometimes when you're trying to determine whether two well-correlated variables are due to cause an effect. The best way to do it is with a controlled experiment, but sometimes you cannot do a controlled experiment. Sometimes it's not feasible. You have to do an observational study due to ethical or practical concerns.

However, it's still possible, though very difficult, to prove cause and effect with a study that isn't an experiment. So let's look at the criteria, and how strict they are.

First, we need consistency. Does the association remain even when other variables are allowed to vary. So does this work across different races, across different genders? All these other things that vary does the association still remain? Do high amounts of the alleged cause lead to high or low amounts of the alleged effect?

Second is sort of like a control. It's not exactly using a control group, But it's sort of what you would do if you had done an experiment. Is the effect absent when the cause is absent? Is the effect present when the cause is present?

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, but we're looking sort of for the same thing. Is the effect present when the cause present, and the effect absent when the cause is absent?

Third, we're looking for correlation. Does as increase in the cause correspond to an increase or, hypothetically, a decrease in the effect? So suppose we're trying to determine whether or not aspirin cures headaches. Does an increase in the amount of aspirin correspond to a decrease in the amount of pain. So an increase, or a decrease in the effect depending on what you're trying to go for.

Fourth. consideration of alternatives. Might there be something else, some lurking variable, that you're missing? That maybe is in the people that are doing this thing, verses some common thread among the people that aren't. So might there be other plausible causes?

And fifth, the connection. What physically might create this effect? What is the physical mechanism behind the effect, and how would it-- how could it plausibly be led to from the cause?

So these are pretty strict requirements. But these are the requirements that we need in order to determine, without an experiment, whether or not two correlated variables are going to be cause and effect related.

So let's go through an example. Consider the claim, "eating a lot of carbohydrates makes you gain weight." So

let's go through these one by one. Is this consistent across different races, different genders? This, more or less, is.

How about control? Is the effect present when the cause is present? Do people who eat lots of carbohydrates gain weight? Well, that's not really the case. You can see a lot of people that eat lots of carbohydrates and don't gain a lot of weight. So I don't-- there's not really a whole mess of reasons to go further, but let's just go through these one by one anyway.

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. So is there anything else besides eating lots of carbohydrates that might make people gain weight? Well, it's possible that people that eat lots of carbohydrates don't exercise as much as people that eat fewer carbohydrates. And so maybe that's what's making them gain weight.

We've considered alternatives and found them to be plausible. So we're going to say that we can't say that this is the only cause.

And then the physical mechanism, is the eating of lots of carbohydrates physically related to weight gain? They are. So this almost passed, and so we didn't really see all of these met. And so we can't say that this claim is 100% true.

Let's do one more. Consider the claim, "smoking causes lung cancer." Consistency. Do we see higher lung cancer rates among smokers across different genders and races? We do. And across different countries. This is true worldwide.

How about control? Now this isn't to say that people can't get lung cancer if they don't smoke. But we see it in much higher rates with people that do smoke, and much lower rates in people that don't smoke. And so we're going to say that that one's met.

Correlation. Are higher-- are people that smoke more-- do those groups of people have higher incidences of lung cancer than people that smoke less? And the answer is yes, they do.

Considering the alternatives, what else might be causing lung cancer? It's possible, I suppose hypothetically, that there's some genetic link they're both causes people to smoke, and predisposes them to lung cancer. But that doesn't really make a whole lot of sense, although it's somewhat plausible. We've considered the alternatives, and can say that this is the most likely cause.

And then the physical connection. Do we understand the physical connection between smoking and lung cancer? We do. We've done experiments using the tar in cigarettes on animals, and those have-- those animals have developed cancerous tumors. So we understand the physical connection.

This passes all of them, and so we can reasonably claim that smoking does cause lung cancer. Now this 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, large contributor.

And we can have different levels in our confidence. In our-- of causation. So let's borrow some terms from the criminal justice system. We have a possible cause, which means you can imagine a scenario where A causes B. So where one thing causes the other.

On possible cause, on those cop shows, it might be where someone becomes a suspect. Probable cause means you're pretty sure that A causes B. Probable cause would be the part in the cop show where the person gets arrested for the crime.

And then cause beyond a reasonable doubt means you cannot think of a scenario where this second variable B, where the response could have been caused by anything other than A. This is the part of the cop show where the person is convicted in a court of law.

And so to recap. The only way to prove 100% definitively causation is with a controlled, randomized experiment. But we have a set of very stringent criteria, whereby we can reasonably conclude that there's a causal link between these two variables based on whether or not they meet these five criteria.

Sometimes the alleged causes don't hold up under the scrutiny, but we can be certain of the ones that do. And so we talked about causation, or causality, those are synonyms, and how we can establish that based on observational studies rather than experiments. Good luck, and we'll see you next time.