

Causation & Experiments

How do we know when something causes something else?

Cause → Effect

Time Order: Cause occurs first in Time, is followed by Effect

Association: When Cause happens, Effect happens (to some degree).

A cause is **sufficient** if it doesn't need anything else to make the effect happen.

A cause is **necessary** if the effect won't happen without it.

A cause can be said to be a **contributing cause** when it is part of a number of causes that lead to the effect.

Do Extracurricular Activities make students into better citizens?

We can go find people who did and did not participate in high school clubs and sports, and see how involved civically involved they are as adults...

...but does this really mean one caused the other? Maybe the same thing that made them into good citizens also made them joiners in high school.

Spuriousness!

The biggest concern when claiming causation.

Maybe something else caused both our “cause”
and “effect”?

Maybe politically engaged parents put their kids
into lots of clubs in high school, train them to be
involved politically as adults too.

Spuriousness

We may think $X \rightarrow Y$

But if $Z \rightarrow X$ and $Z \rightarrow Y$, it may look like $X \rightarrow Y$ when really it doesn't at all.

We need to know about **every** possible Z . Or at least somehow prevent every possible Z from happening...

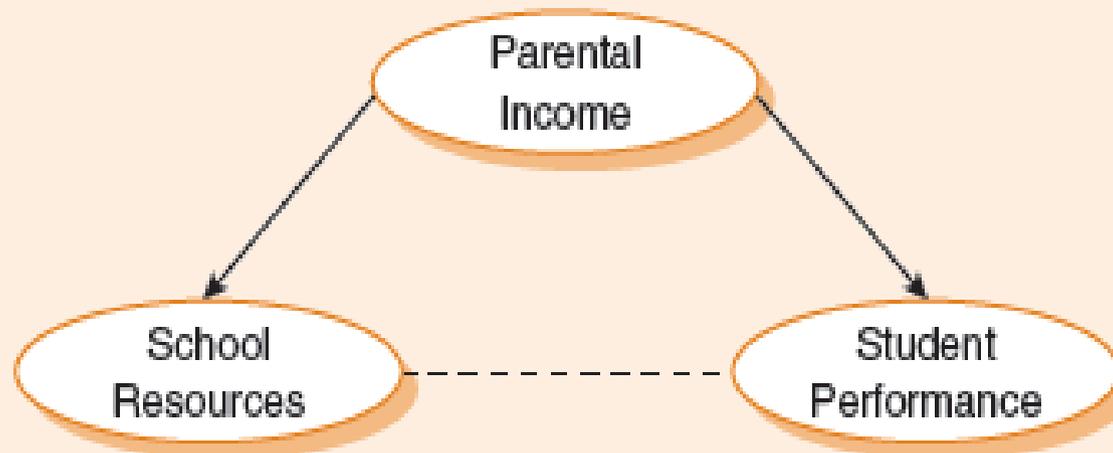
EXHIBIT 6.1

A Spurious Relationship Revealed

School resources are associated with student performance; apparently, a causal relation.



But in fact, parental income (a third variable) influences both school resources and student performance, creating the association.



Experiments

Experiments are designed to eliminate spuriousness, isolate causation.

The focus is on control:
Making everything the same for each subject,
except for what we think the cause is.

Two comparison groups

Treatment Group

- Gets the proposed cause (the treatment)

The difference between the effect (the outcome of interest) in the two groups should be entirely due to the cause...

Control Group

- Doesn't get the cause (no treatment), but otherwise is treated exactly the same.

... if the two groups are truly equal in all other respects (ceteris paribus).

Equal Comparison Groups through The Magic of Random Assignment

Just as random sampling is the magic tool to give us representative samples from a whole...

...random assignment can give us equivalent groups for comparison.

Random sampling is for generalizability
Random assignment is for internal (causal)
validity

For instance, let's say we take 100 college students, and test their IQs. Then we randomly pick 50 of them to be frightened by a clown, while the other get an uneventful 10 minute break. Then we re-test them all.

The difference in the change of their IQ scores between the clowned and unclowned group should be entirely due to the clown effect.

But what if the clown group has more easily-frightened people in it?

That's very unlikely, since we randomly selected them. We shouldn't have to worry about any difference between the groups (any possible Zs) since only luck chose them to be scared.

Example: Asian American-ness Experiment

Researchers approached Asian American students on a college campus.

Randomly-selected 50% were asked what country they were from.

They were then surveyed about attitudes towards America and knowledge of Americana.

Natural Experiments

Sometimes we can't use experiments, for practical or ethical reasons.

For instance, how does the quality of a high school affect student achievement?

Researchers try to find real world situations that mimic random selection.

Chicago Public School Lottery

Steven Levitt used the fact that students in Chicago were randomly assigned to high achievement and low achievement schools as a real-world experiment.

Both the treatment and control students had parents who wanted them in the better schools.

At the end of high school, there was no difference in their test scores.

Quasi Experiments

When true and natural experiments aren't possible, researchers can try to make equivalent treatment and control groups through means other than random selection.

This usually doesn't work as well.

Matching

Attempting to make equivalent comparison groups by matching cases on all important characteristics.

But you need to know every single factor that could possibly affect the outcome variable, and match on that.

Threats to Causal Validity

Noncomparable Groups

Endogenous Change

History (External Events)

Contamination

Treatment Misidentification

Noncomparable Groups

Selection Bias
(such as self selection)

Differential Attrition
Those who drop out are different from those who
stay

Endogenous Change

Subject change over time naturally.

Regression to the Mean – extreme measures tend to be unusual for the subjects more than indication of unusual subjects.

Subjects chosen for their extremeness tend to measure closer to the average the next time.

History (External Events)

Imagine if you were running an long-term experiment that measured perceptions of world events, and 9/11 happened in the middle of it.

Or more simply, you're running a lab experiment, but on some days thunder and lightning distract your subjects.

Contamination

The treatment group somehow “contaminates” the control group, like two petri dishes in a biology lab that are left too close together.

Even knowing that another group is being treated better/worse may affect subjects.

The John Henry Effect – Competition makes people try harder.

Its best to keep the groups completely separate and ignorant of each other.

Treatment Misidentification

Something mattered, but not what you think...

Placebo Effect: you think a drug will work, so it does

Hawthorne Effect: observed subjects behave differently than they normally would (for instance, they work harder)

Experimenter Expectation: researchers treat the treatment group subtly different in other ways. They expect or want them to be different from the control, which can subtly influence the outcome.
(Avoid this with **double-blind** procedures)

Vignette Studies

Give the subjects little stories to read, then ask them questions.

For instance, you could give people a detailed story about a family on welfare, and ask people questions to gauge sympathy. But some randomly read about a white family, some randomly read about a hispanic family, etc. Or some read about a two parent household, some read about a single mother.

Evaluating\Grading Experiments

Shelly Correll had college students judge the resumes of female job applicants, some of which were clearly mothers, others not. The mothers were considered less promising candidates.

Darley & Gross had subjects grade tests of elementary school students, some of which were identified as low income, others upper middle class. The lower income students were graded more harshly.

Devah Pager's Audit Study of Racial & Felony Discrimination

The Treatments: Race and Felon status were randomly varied.

The Results: Since everything else about the applicants was the same, and the assignment to treatments was equal, this was strong evidence that the cause of the much lower call-back rate for black applicants had to be about race.

Group Task Experiments

Another tradition in Sociology looks at how small groups accomplish tasks, and the hierarchies that develop within them.

Cecilia Ridgeway had subjects work in small groups to perform a task, but one of the members was a **confederate** who broke the procedural rules they established. When the confederate was male (and especially when the task was defined as mathematical), the subjects were more likely to follow the confederate.

Quasi Experiments

When the assignment to treatment and control isn't really random.

Before and After Tests: subjects pretest is effectively the control group, their posttest is the after-treatment group. They are assumed to be equivalent to themselves otherwise...

An Ex Post Facto Control Group

Using existing data about high school students, we can match them to compare students who do and don't participate in extracurriculars.

Only compare students with the exact same of everything that could possibly matter to voting and club/sport participation.

The data also talks to them as adults, and asks if they vote.

If we designed this ahead of time to do this, it would be a Quasi Experiment. Because we didn't and are using existing data in this way, its not.

Hypothetical Example:

During WWII, soldiers drew straws to see who would go on a dangerous mission. Those who didn't never saw combat, but those who did go saw a lot of it.

Researchers later track them down to determine the effects of combat on long term mental health.

Another hypothetical:

Suppose that when you login in to Facebook, you have a 50% chance of being shown a picture of a smiling person, and a 50% chance of seeing a picture of an expressionless person. Facebook then tracks your activity to see if you are more or less likely to “like” comments and photos for that session.