

5.2: Studies of Causality

If we want to draw conclusions about *causality*, observations are insufficient. This is because simply seeing *B* always follow *A* out in the world does not tell us that *A* causes *B*. For example, maybe they are both caused by *Z*, which we didn't notice had always happened before those *A* and *B*, and *A* is simply a bit faster than *B*, so it seems always to proceed, even to cause, *B*. If, on the other hand, we go out in the world and do *A* and then always see *B*, we would have more convincing evidence that *A* causes *B*.

Therefore, we distinguish two types of statistical studies

DEFINITION 5.2.1. An **observational study** is any statistical study in which the researchers merely look at (measure, talk to, *etc.*) the individuals in which they are interested. If, instead, the researchers also change something in the environment of their test subjects before (and possibly after and during) taking their measurements, then the study is an **experiment**.

EXAMPLE 5.2.2. A simple survey of, for example, opinions of voters about political candidates, is an observational study. If, as is sometimes done, the subject is told something like “let me read you a statement about these candidates and then ask you your opinion again” [this is an example of something called **push-polling**], then the study has become an experiment.

Note that to be considered an experiment, it is not necessary that the study use principles of good experimental design, such as those described in this chapter, merely that the researchers *do something* to their subjects.

EXAMPLE 5.2.3. If I slap my brother, notice him yelp with pain, and triumphantly turn to you and say “See, slapping hurts!” then I’ve done an experiment, simply because I *did something*, even if it is a stupid experiment [tiny non-random sample, no comparison, *etc.*, *etc.*].

If I watch you slap someone, who cries out with pain, and then I make the same triumphant announcement, then I’ve only done an observational study, since the action taken was not by me, the “researcher.”

When we do an experiment, we typically impose our intentional change on a number of test subjects. In this case, no matter the subject of inquiry, we steal a word from the medical community:

DEFINITION 5.2.4. The thing we do to the test subjects in an experiment is called the **treatment**.

Control Groups

If we are doing an experiment to try to understand something in the world, we should not simply do the interesting new treatment to all of our subjects and see what happens. In a certain sense, if we did that, we would simply be changing the whole world (at least the world of all of our test subjects) and then doing an observational study, which, as we have said, can provide only weak evidence of causality. To really do an experiment, we must *compare* two treatments.

Therefore any real experiment involves at least two groups.

DEFINITION 5.2.5. In an experiment, the collection of test subjects which gets the new, interesting treatment is called the **experimental group**, while the remaining subjects, who get some other treatment such as simply the past common practice, are collectively called the **control group**.

When we have to put test subjects into one of these two groups, it is very important to use a selection method which has no bias. The only way to be sure of this is [as discussed before] to use a random assignment of subjects to the experimental or control group.

Human-Subject Experiments: The Placebo Effect

Humans are particularly hard to study, because their awareness of their environments can have surprising effects on what they do and even what happens, physically, to their bodies. This is not because people fake the results: there can be real changes in patients' bodies even when you give them a medicine which is not physiologically effective, and real changes in their performance on tests or in athletic events when you merely convince them that they will do better, *etc.*

DEFINITION 5.2.6. A beneficial consequence of some treatment which should not directly [*e.g.*, physiologically] cause an improvement is called the **Placebo Effect**. Such a “fake” treatment, which looks real but has no actual physiological effect, is called a **placebo**.

Note that even though the Placebo Effect is based on giving subjects a “fake” treatment, the effect itself *is not fake*. It is due to a complex mind-body connection which really does change the concrete, objectively measurable situation of the test subjects.

In the early days of research into the Placebo Effect, the pill that doctors would give as a placebo would look like other pills, but would be made just of sugar (glucose), which (in those quite small quantities) has essentially no physiological consequences and so is a sort of neutral dummy pill. We still often call medical placebos **sugar pills** even though now they are often made of some even more neutral material, like the starch binder which is used as a matrix containing the active ingredient in regular pills – but without any active ingredient.

Since the Placebo Effect is a real phenomenon with actual, measurable consequences, when making an experimental design and choosing the new treatment and the treatment for the control group, it is important to give the control group *something*. If they get nothing, they do not have the beneficial consequences of the Placebo Effect, so they will not have as good measurements as the experimental group, even if the experimental treatment had no actual useful effect. So we have to equalize for both groups the benefit provided by the Placebo Effect, and give them both an treatment which looks about the same (compare pills to pills, injections to injections, operations to operations, three-hour study sessions in one format to three-hour sessions in another format, *etc.*) to the subjects.

DEFINITION 5.2.7. An experiment in which there is a treatment group and a control group, which control group is given a convincing placebo, is said to be **placebo-controlled**.

Blinding

We need one last fundamental tool in experimental design, that of keeping subjects and experimenters ignorant of which subject is getting which treatment, experimental or control. If the test subjects are aware of into which group they have been put, that mind-body connection which causes the Placebo Effect may cause a systematic difference in their outcomes: this would be the very definition of bias. So we don't tell the patients, and make sure that their control treatment looks just like the real experimental one.

It also could be a problem if the experimenter knew who was getting which treatment. Perhaps if the experimenter knew a subject was only getting the placebo, they would be more compassionate or, alternatively, more dismissive. In either case, the systematically different atmosphere for that group of subjects would again be a possible cause of bias.

Of course, when we say that the experimenter doesn't know which treatment a particular patient is getting, we mean that they do not know that at the time of the treatment. Records must be kept somewhere, and at the end of the experiment, the data is divided between control and experimental groups to see which was effective.

DEFINITION 5.2.8. When one party is kept ignorant of the treatment being administered in an experiment, we say that the information has been **blinded**. If neither subjects nor experimenters know who gets which treatment until the end of the experiment (when both must be told, one out of fairness, and one to learn something from the data that was collected), we say that the experiment was **double-blind**.

Combining it all: RCTs

This, then is the gold standard for experimental design: to get reliable, unbiased experimental data which can provide evidence of causality, the design must be as follows:

DEFINITION 5.2.9. An experiment which is

- *randomized*
- *placebo-controlled*.
- *double-blind*

is called, for short, a **randomized, controlled trial [RCT]** (where the “placebo-” and “double-blind” are assumed even if not stated).

Confounded Lurking Variables

A couple of last terms in this subject are quite poetic but also very important.

DEFINITION 5.2.10. A **lurking variable** is a variable which the experimenter did not put into their investigation.

So a lurking variable is exactly the thing experimenters most fear: something they didn't think of, which might or might not affect the study they are doing.

Next is a situation which also could cause problems for learning from experiments.

DEFINITION 5.2.11. Two variables are **confounded** when we cannot statistically distinguish their effects on the results of our experiments.

When we are studying something by collecting data and doing statistics, confounded variables are a big problem, because we do not know which of them is the real cause of the phenomenon we are investigating: they are statistically indistinguishable.

The combination of the two above terms is the worst thing for a research project: what if there is a lurking variable (one you didn't think to investigate) which is confounded with the variable you did study? This would be bad, because then your conclusions would apply equally well (since the variables are statistically identical in their consequences) to that thing you didn't think of ... so your results could well be completely misunderstanding cause and effect.

The problem of confounding with lurking variables is particularly bad with observational studies. In an experiment, you can intentionally choose your subjects very randomly, which means that any lurking variables should be randomly distributed with respect to any lurking variables – but controlled with respect to the variables you are studying – so if the study finds a causal relationship in your study variables, it cannot be confounded with a lurking variable.

EXAMPLE 5.2.12. Suppose you want to investigate whether fancy new athletic shoes make runners faster. If you just do an observational study, you might find that those athletes with the new shoes do run faster. But a lurking variable here could be how rich the athletes are, and perhaps if you looked at rich and poor athletes they would have the same relationship to slow and fast times as the new- vs old-shoe wearing athletes. Essentially, the variable *what kind of shoe is the athlete wearing* (categorical with the two values *new* and *old*) is being confounded with the lurking variable. So the conclusion about causality might be false, and instead the real truth might be *wealthy athletes, who have lots of support, good coaches, good nutrition, and time to devote to their sport, run faster*.

If, instead, we did an experiment, we would not have this problem. We would select athletes at random – so some would be wealthy and some not – and give half of them (the experimental group) the fancy new shoes and the other half (the control group) the old type. If the type of shoe was the real cause of fast running, we would see that in our experimental outcome. If really it is the lurking variable of the athlete's wealth which matters, then we would see neither group would do better than the other, since they both have a mixture of wealthy and poor athletes. If the type of shoe really is the cause of fast running, then we would see a difference between the two groups, even though there were rich and poor athletes in both groups, since only one group had the fancy new shoes.

In short, experiments are better at giving evidence for causality than observational studies in large part because an experiment which finds a causal relationship between two variables cannot be confounding the causal variable under study with a lurking variable.

This page titled [5.2: Studies of Causality](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Jonathan A. Poritz](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.