PSY 225 – Research Methods
Professor Gernsbacher's Lecture Video "How to Evaluate Alternative Research Hypotheses"

In this video, you're going to learn how to evaluate alternative research hypotheses.

CLICK: You've previously learned how the terms Reliability and Validity are used in research. Reliability refers to how reliable, meaning how repeatable and reproducible, a research phenomenon is.

Validity refer to how valid, meaning how justifiable, a research phenomenon is. When talking about validity, many researchers distinguish between

CLICK: External Validity and Internal Validity.

External Validity refers to whether the study can be justified externally. For example, whether the study's results and conclusions can generalize beyond that study's population to other populations or beyond that study's operational definitions to other operational definitions.

Internal Validity refers to whether the study can be justified internally. Questions about a study's internal validity are often called the scary word "Threats" as in

CLICK: "Threats to Internal Validity."

I prefer the less scary term, candidates for

CLICK: Alternative Hypotheses, so that's the term we'll be using here, as we learn how to evaluate alternative research hypotheses -- because evaluating alternative hypotheses is the same process as evaluating threats to a study's internal validity.

And in case, all of this jargon sounds overwhelming, don't fear. You already know a couple of ways to evaluate alternative research hypotheses. So, let's start with an alternative research hypothesis, also known as a threat to internal validity, that you already know:

CLICK: **Correlation Isn't Causation**. You've no doubt learned that just because two variables are correlated, that doesn't mean that one variable caused the other. For example, just because the amount of ice cream that is sold each month correlates with the number of drownings that are reported each month, that correlation does not mean that eating ice cream causes drowning.

Two variables might be correlated, but another variable might cause both. For example, in the ice-cream and and drownings correlation, the variable most likely causing both ice cream sales and reported drownings to increase or decrease is the season of the year. More ice cream is sold in the summer months than the winter months, and more drownings occur in the summer months than the other winter months.

So, a third variable, season of the year causes both primary variables, ice cream sales and drownings, to increase and decrease at the same rate, but neither of the two primary variables cause each other.

You've also no doubt learned that just because two variables are correlated, for example, skipping breakfast and being over-weight, doesn't mean that the first variable causes the second. The second variable could cause the first, for example, people might skip breakfast because they are over-weight.

And two variables might be correlated simply because of coincidence, for example,

CLICK: the number of people who drowned per year, rather than per month, by falling into a swimming pool, and the number of films Nicholas Cage appeared in each year. These two variables are reasonably highly correlated. For example, in 2003, there were relatively fewer drownings and fewer films that Nicholas Cage appeared in, and in 2007, there were relatively more drownings and more films that Nicholas Cage appeared in. But most likely drownings don't cause Nicholas Cage to appear in films, and Nicholas Cage appearing in films doesn’t cause drowning. And because there’s not an obvious third variable causing both, it’s likely that the correlation is coincidence. Thus,

CLICK: **Correlation Isn’t Causation** is a viable alternative hypothesis whenever a study claims that one variable CAUSED another, but all the study has demonstrated is that one variable is CORRELATED with another.

Let’s look at another alternative hypothesis that you might already be familiar with.

CLICK: **Sampling Bias or Participant-Selection Bias.**

You know that a research sample should be unbiased. It should be representative of whichever target population it is intended to represent. However, sometimes research studies fail to adequately sample their target populations and, in so doing, they create a sampling or participant-selection bias.

Often the problems of sampling bias or participant-selection bias arise because the sample isn’t randomly selected but is instead based on participants either volunteering for the study or being selected for a study. For example, if we wanted to study the effects of meditation on college performance, and we asked for volunteers for our participants, we are likely to obtain a biased sample -- if mostly college students who are already experienced meditators volunteered for our study. And if that happens, we’d have a sampling bias.

As another example, if I, as an instructor, wanted to study the effects of a new teaching technique, but I selected for my sample of participants only certain students I knew, I might be creating a participant-selection bias. Maybe I sub-consciously chose students who I knew would benefit from the new teaching technique. If so, I’d have a participant-selection bias.

Sampling Biases and Participant-Selection Biases are particularly problematic in studies that aim to compare two samples. If one sample is biased, the validity of the comparison to the other sample is threatened, which is why we consider Sampling Biases and Participant-Selection Biases threats to internal validity.

A related threat to internal validity, or as I like to say, a candidate for an alternative hypothesis, is

CLICK: **Researcher Bias.**

Researcher bias is like participant-selection bias because it’s caused by the researcher. But researcher bias doesn’t occur during participant selection. Rather researcher bias occurs during other aspects of the study, particularly those aspects that involve the researcher interacting with the participants.

For example, let’s say that a researcher has a hypothesis that women will respond differently on a questionnaire than will men. That researcher might sub-consciously treat the women participants in their study differently before or while they are responding to the questionnaire than the researcher treats the men participants.

As another example, remember the example I used before, the one in which I, as an instructor, wanted to study the effects of a new teaching technique? As I mentioned before, if I selected my sample of participants in a biased way that would be a participant-selection bias. But let’s say I selected my participants completely randomly. No participant-selection bias at all.

However, if once the participants began the study, I, the researcher, was also the person who interacted with the participants, I might sub-consciously interact with the participants using the new teaching technique more enthusiastically than the participants using the old teaching technique. That would be a researcher bias.

Some effects of researcher bias can be quite subtle but can still threaten internal validity in a big way. Researcher bias is such a powerful threat to internal validity, that many studies use what’s known as a double-blind technique.
CLICK: In a “Double-Blind” study, or what I prefer to call a

CLICK: "Double-Naïve" study because "naïve" is both more accurate and less ableist, both the

CLICK: researchers who interact with the participants AND

CLICK: the participants are kept naïve about the study's hypothesis and goals. That way, the researcher can't subtlety affect the outcome of the study because the researcher who interacts with the participants doesn't know the study's preferred outcome. Keeping the researcher or researchers who interact with the participants naïve, as in a Double-Naïve study, is the primary way to reduce Researcher bias.

In fact, in Double-Naïve treatment studies, neither the researchers nor the participants know whether who is in the treatment group versus who is in the control group.

Double-Naïve studies contrast with what's known as Single-Blind or what I call,

CLICK: Single-Naïve studies, in which only the research participants are kept naïve but the researchers are not kept naïve.

However, even in both double-naïve and single-naïve studies, there can be threats to validity that are due to

CLICK: **Participant Bias**.

And that's because just the very fact that participants know that they are in a research study can often cause them to behave differently, which is captured by the term

CLICK: Hawthorne Effect

In the 1920s, the owners of the

CLICK: Hawthorne telephone-equipment factory in Cicero, Illinois, wanted to study whether improving the factory's interior lighting and giving factory workers better scheduled breaks would improve their productivity.

The workers did increase their productivity, but later it was discovered that the improved productivity was primarily due to the workers, who were the participants in this study, being aware of being in a research study. Their productivity improved because of the attention they received from being participants in the study, not because of the manipulations of the study, which were things like the improved lighting and break schedule.

To this day, we refer to a Hawthorne effect whenever participants respond differently simply because they know that they are in a research study. We refer to this type of reaction as

CLICK: reactivity, and if you are interested in reading more about reactivity, or the Hawthorne Effect, I encourage you to Google it.

The participants in the Hawthorne factory were observed by researchers, but imagine the threats to validity that could occur if instead the participants had provided self-reports of their productivity – or their activities, behaviors, or attitudes, which leads us to another alternative hypothesis, known as the

CLICK: **Self-Report Bias**.

The self-report bias is exactly what its name implies; it's the threat to a study's internal validity caused by the data being based on participants' self-report. Although for some phenomena, self-report is one of the few ways to collect data, self-report can be complicated by the participants' desire to come off looking good in a study or even bad.

Self-report biases are particularly problematic in treatment studies because participants tend to report that the treatment was effective, given the amount of time or effort it involved.

There are a few more candidates for alternative hypotheses to cover, but let me make sure that we're clear on the similarly worded ones we've covered so far.

CLICK: Sampling Bias or Participant-Selection Bias occurs when the sample of participants are chosen in some biased way, and those sampling or participant-selection biases could affect the study's results.

CLICK: Researcher Bias occurs when the researcher isn't naïve about the study's hypotheses and the researcher could subtlety bias the participants to respond in ways that confirm the study's hypothesis.

CLICK: Participant Bias occurs when the participants are reactive, that is, they respond differently simply because they are aware of being in a research study.

CLICK: Self-report bias is a form of Participant Bias, but it's specific to measures that are taken through participants' self-report.

OK, let's move on to three other effects that can pose threats to a study's internal validity and therefore provide an alternative hypothesis for the study's results.

CLICK: The "**Effect of History**" is a research term for an effect that's rather intuitive, despite the awkward name. In studies that measure phenomena twice, say a pre-test and a post-test, there are several things that could occur in between those two test points that might affect the results.

For example, probably every pre- versus post-test psychological science study conducted in late summer through mid-fall of the year 2001 was confounded by the occurrence of 9/11 on September 11th. Unless the study was specifically designed to measure the effects of a national crisis, the study was undoubtedly complicated by that horrific event's occurrence.

Threats to internal validity that fall under the umbrella of Effects of History don't have to be as catastrophic as 9/11. Even something as simple as another person's cell phone unexpectedly ringing during a fifteen-minute experiment can complicate the results.

A specific Effect of History is called

CLICK: the "**Maturation Effect**," which refers specifically to the maturation or development that any organism will experience between one testing point to another. For example, if students are tested at the beginning of the semester and the end of the semester, they might have developed during that semester in ways that complicate the study.

Conversely related to the Maturation Effect is the

CLICK: the "**Mortality Effect**," which refers specifically to organisms, not necessarily dying, although that can happen, but no longer being included in the participant sample by the end of a study.

For example, in a study that examines a new teaching approach in a college course, some students might drop the course or even drop out of college completely between the pre-test and the post-test. And these students who dropped or dropped out might be the students for whom the new teaching approach worked well or didn't work at all – but we don't know, because of the Mortality Effect.

As my example suggested, the Mortality Effect is particularly problematic in treatment studies, because a study might show a positive effect of the treatment simply because the participants for whom the treatment didn't work dropped out of the study. Or a study might show a null effect of a treatment because the participants for whom the treatment worked well dropped out – thinking that they didn't need to stay in the study because they'd already benefited from the treatment's effect.

So, the Mortality Effect can be an alternative hypothesis for positive findings, for negative findings, or for null findings. In fact, the Mortality Effect is a type of sampling bias that occurs during a study. So, just as sampling bias can be an alternative hypothesis for positive, negative, or null findings, so can the Mortality Effect.

Related to the Maturation Effect is the contrast between a

CLICK: **Cross-Sectional versus a Longitudinal Effect**. Cross-Sectional and Longitudinal are two names for two different research designs used to measure temporally caused change. In a cross-sectional design, participants of different levels of development are all studied at one point in time; in a longitudinal design, participants are studied across time, as they progress through different levels of development.

For example, let's say we hypothesized that college improves students' writing ability.

CLICK: Using a cross-sectional design, we would gather a sample of

CLICK: Freshman, sophomores, juniors, and seniors, test them all at one point in time and see if the seniors' writing ability is better than the juniors, which is better than the sophomores, and then the freshman. Our design is cross-sectional because we are studying participants of different developmental levels at the same point in time.

Conversely, using

CLICK: a longitudinal design, we would gather a sample of only

CLICK: freshman, and then test these freshman on their writing skills not only during their freshman year, but also during their

CLICK: sophomore year, their junior year, and their senior year. This design is longitudinal because we are studying the same participants longitudinally, which means across different developmental levels at different points in time.

Using either a cross-sectional or a longitudinal design, we can examine development. But if we're using a cross-sectional design, and we want to make strong claims about age or other markers of time causing development, we must first consider an alternative hypothesis: Maybe the participants that we tested all at the same time actually differed in ways beyond our hypothesis.

For example, in our cross-sectional study of writing ability, maybe, some other reason than the one we hypothesized, led to the seniors being different from the freshman, which led us to believe that it was four-years of college that improved their writing skills, when it wasn't.

All of which is to say that stronger claims of causality can be made from longitudinal designs than from cross-sectional designs. On the other hand, cross-sectional designs avoid the problems of the Maturation Effect and the Mortality Effect that can hamper the results of longitudinal designs.

Lastly, we get to my absolute favorite alternative hypothesis, and it is

CLICK: **Regression to the Mean**. Although the term regression is an inferential statistics technique, the term "regression to the mean" has its own meaning. Regression to the mean occurs when participants are selected for having extreme attributes, for example, children who are all in the lowest percentile for reading ability, or adults who are all in the highest percentile for depression.

Typically, what happens is that with just the passage of time or with repeated testing, these extreme scores become a bit more typical, which is why we call this phenomenon "regression to the mean."

For example, let's say that we

CLICK: tossed a quarter six times. We might find that it comes up heads three times and it comes up tails three times.

Let's say we tossed that same quarter another six times.

CLICK: We might find that it comes up tails six straight times. Six straight tails is still within the realm of randomness – we don't have a trick quarter. But six straight tails is somewhat of an extreme observation.

So, we toss the same coin again six times,

CLICK: and this time it comes up tails only four times. We could say that our third toss of the same coin demonstrates regression to the mean – the array of heads versus tails is more typical.

In the same way, regression to the mean is often a function of randomness in a test result or fluctuations in performance, mood, or behavior. When a child who has tested extremely low on a reading test is tested again, the second (or third) test might be closer to the child's actual ability. While that child's actual ability might not be the average of all other children, a second or third test should produce test results that are more in line with the average of that child's actual ability.

Indeed, one of my favorite examples of regression to the mean representing actual ability, not necessarily average ability, but actual ability is the so-called

CLICK: *Sports Illustrated* curse. Those of you who are sports fans know that there's supposedly a curse associated with being on the cover of *Sports Illustrated* magazine. That's because after some teams or athletes have been featured on the cover of *Sports Illustrated*, they've performed relatively more poorly, which has given rise to the suspicion that being on the cover of *Sports Illustrated* is a curse!

But the supposed *Sports Illustrated* curse could simply be due to regression to the mean. When do teams or athletes get featured on the cover of *Sports Illustrated*? It's when they've demonstrated an extremely high, and often unexpectedly high, level of performance – for example, after the Broncos went 4-0 in 2009. And regression to the mean predicts that most extreme values will eventually become a bit more typical – for that individual or team.

CLICK: Regression to the mean can also explain the improvement shown after treatment of particularly extreme phenomena. For example, some studies show that extremely depressed patients improve even if in they are in control group because their extreme depression regresses to more typical levels of depression – typical for that person.

That is, a typical level of depression for a person with chronic depression will still be higher than a typical level of depression for a person without chronic depression, just like the Broncos' typical level of winning might still be higher than another team. But the notion is that usually very extreme scores or very extreme performance regresses to more typical levels – typical of that person. And that's what regression to the mean means.

OK, so we've now covered nine threats to internal validity or what I've called nine candidates for alternative hypotheses.

CLICK: Correlation Isn't Causation; Sampling Bias or Participant-Selection Bias; Researcher Bias; Participant Bias; and Self-Report Bias;

CLICK: Effect of History and its special case, the Maturation Effect; the Mortality Effect; the Cross-Sectional versus Longitudinal Effect; and Regression to the Mean.

Knowing these nine candidates for alternative hypotheses should equip you well to evaluate psychological science.