

Chapter 5: Experiments, Good and Bad

- Recall: In an **experiment**, a treatment is actively *imposed* on the cases. The goal of an experiment is to gain information regarding the effects (if any) of the treatments on the response of interest.
- The specific experimental conditions are called **treatments**. Often a treatment is composed of several **factors**. For example, a drug *treatment* may consist of a combination of drugs. Each drug in the treatment is considered a *factor*.
- In an experiment, the **experimental units** are the objects subjected to specific experimental treatments. If the experimental units are humans, they are called **subjects** or **participants**.
- A **response variable (dependent variable)** is a variable that measures an outcome or result of interest in a study. An **explanatory variable (independent variable)** is a variable that we believe explains or causes changes in the response variable.
- Example: A mason wants to study the effect of three specific firing temperatures on the density of bricks. 24 bricks are randomly selected, and 8 bricks are randomly assigned to each firing temperature. The brick densities are measured one hour after each brick is fired.
 1. What is the response?
 2. What are the factor(s)?
 3. How many treatments are there?
 4. What are the experimental units?
 5. How many experimental units are there?
- Example: An aluminum alloy manufacturer produces grain refiners. The company produces the product in four furnaces. Each furnace is known to have its own unique operating characteristics. The process engineers want to know whether the stirring rate affects the grain size of the product. Each furnace was run at four stirring rates (5, 10, 15, and 20 rpm) with the order of these stirring rates randomized for each furnace. Grain size data was collected for each furnace and stirring rate combination.
 1. What is the response?
 2. What are the factor(s)?
 3. How many treatments are there?
 4. What are the experimental units?
 5. How many experimental units are there?
- In a study, **confounding** occurs if the researcher cannot determine the effects of specific factors on the response because the factor effects are mixed and cannot be separated. Lurking variables (i.e., variables that have an important effect on the response variable but is not one of the variables of interest in the study) can often cause confounding.
 - In the early 1990s, Pfizer Corporation developed fluconazole, a drug to prevent fungal infections in hospital patients. Several studies found the new drug to be more effective than the commonly-used drug called amphotericin B. A subsequent analysis by other

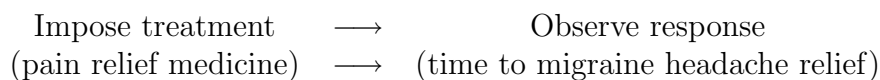
researchers, however, found that the older drug had been administered orally (in these studies) when it was supposed to be given by injection. This introduced a source of confounding into the studies. That is, the original researchers thought the results showed the new drug to be more effective than the older drug, but they had not taken into account the confounding effects of how the drug was administered. Once the new researchers took this effect into account, they found that the new drug was no more effective than the older drug.

- Example: Suppose an engineer wants to determine if there is any effect of increasing production temperature on the breaking strength of an industrial fiber (e.g, Kevlar which is used to make bullet-proof vests). The engineer tests the breaking strength at 150°, 250°, and 350°, and collected the following data.

Temperature	Strength (lbs)			Temperature	Operator	Strength (lbs)		
150°	145	147	142	150°	1	145	147	142
250°	133	130	133	250°	2	133	130	133
350°	126	120	122	350°	3	126	120	122

- From the table on the left, what can we conclude about the effect of temperature on the breaking strength of this fiber (*assuming conditions remain constant throughout the experiment except for temperature changes*)?
- From the table on the right, what can we conclude about the effect of temperature on the breaking strength of this fiber (*assuming conditions remain constant throughout the experiment except for temperature and operator changes*)?

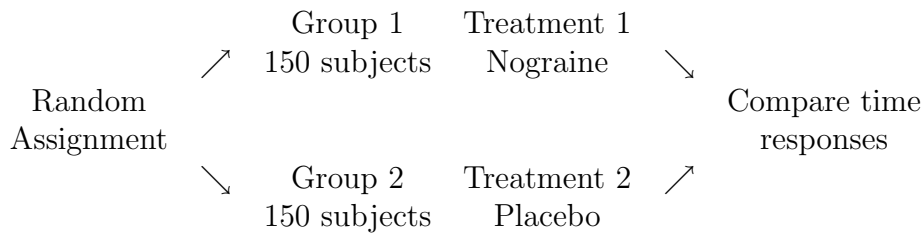
- Recall: Conclusions regarding causality are tentative from observational studies because uncontrolled (lurking) variables may be confounded with the explanatory factors. Conclusions regarding causality can be made with more confidence when a properly designed experiment is used.
- Some benefits of a designed experiment over an observational study are an experiment allows for (i) the study of effects of factors that are of particular interest, (ii) the control of factors not of interest, and (iii) the study of combined effects of several factors simultaneously.
- Discussion: (One-track) experiment: Data was collected from 300 subjects suffering from migraine headaches. The goal was to determine if the drug Nograine was effective in relieving migraine headache symptoms after onset of a migraine headache. Each subject received the same 500mg dosage of Nograine.



What is the major flaw with this design?

- In a **comparative experiment**, two or more treatments are applied to the set of experimental units. The responses of experimental units having different treatments are then compared.
- When dealing with human or animal subjects, control, the experiment often has *treatment* and *control* groups. **Treatment groups** are groups of subjects who receive an experimental treatment. A **control group** is a group of subjects that do not receive an experimental treatment.

- It is common for a control group to receive a placebo. A **placebo** lacks any active ingredients of a treatment that is being studied but is identical in appearance to the treatment. Therefore, study participants cannot distinguish the placebo from the real treatment.
- A **placebo effect** corresponds to the situation when study participants respond favorably to a placebo treatment because they believe they are receiving a useful treatment. The responses of units in a treatment group are compared to the responses of units in the control group to see if any favorable responses in a treatment group can be attributed to the treatment instead of a placebo effect.
- Suppose we modify the one-track experiment so that 150 subjects are randomly assigned to Group 1 and will receive Nograine, while the other 150 subjects are assigned to Group 2 and will receive a placebo.



Now we can assess whether or not Nograine is effective at reducing time to migraine relief.

- Read “Is it or isn’t it a Placebo?” on page 97 of C&C.
- Experiments that study the effectiveness of medical treatments on actual patients are called **clinical trials**.
- See “The Salk Polio Vaccine” on Handout #3.
- Comparative experiments (i) are more successful than uncontrolled (that is, no comparison or control group) or observational studies and (ii) reduce the possibility of confounding treatment effects. The following are examples of comparative experiments.

Three Principles of Experimental Design

Principle 1: **The Need to Control** the effects of factors that are not of interest. These can be environmental factors (like temperature, lighting, noise levels), psychological factors (like the temperament of the researcher, placebo effects), etc.

- When the goal is to compare the effectiveness of a new treatment relative an existing standard treatment or commonly-used alternative, the control group is often a group that receives a the standard or currently-used or treatment. For example, if you want to market a new pain-relief drug, you would want to perform an experiment that provides evidence that the new drug is more effective than a product like Tylenol for relieving pain. In this case, the treatment group receives the new drug and the control group would receive the Tylenol.
- Another potential factor to control is an *experimenter effect*. An **experimenter effect** occurs when a researcher somehow influences subjects through such factors as body language, tone of voice, attitude, etc.
- To reduce the effects of factors not of interest, the following experimental protocol (called **blinding**, is often used.

- An experiment is said to be **single blind** if the participants do not know whether they are members of a treatment or a control group, but the experimenters do know (or vice versa).
 - An experiment is said to be **double blind** if neither the participants nor the experimenters know not know which participants are members of the treatment and the control groups. This requires a third-party to keep track of the assignments of subjects to treatments.
- The goal of control is to prevent factors not of interest to bias responses of the subjects.
 - The Salk Polio Vaccine study was actually a double-blind study because neither the participants (the children) nor the experimenters (the doctors and nurses administering the injections and diagnosing polio) knew who got the real vaccine and who got the placebo.
 - See the “General Adverse Events for Seldane D” example on Handout #3.

Principle 2: Random Assignment is the process of randomly assigning subjects to treatment and control groups to create groups that are similar (except for chance variation). Random assignment reduces the chance of systematic differences and reduces the chance of confounding the treatment effects.

Principle 3: Replication is the process of repeating treatments on a sufficient number of subjects.

- Replication can reduce the effects of chance variation. This will make the experiment more sensitive for detecting systematic effects of the treatments. For example, you would not want to draw a conclusion based on the reaction of 1 person to each treatment.

Discussion: For each of the following experiments, identify any problems and explain how the problems could have been avoided?

1. A new drug for attention deficit disorder (ADD) is supposed to make affected children more polite. Randomly selected children suffering from ADD are divided into treatment and control groups. The experiment is single-blind. Experimenters evaluate how polite the children are during one-on-one interviews.
2. Researchers wonder if the effects of a rare degenerative disease can be slowed by exercise. They identify six people suffering from the disease, and randomly assign three to a treatment group that exercises every day and three to a control group that avoids exercise. After six months, they compare the amounts of degeneration in each group.
3. Education researchers wonder if listening to classical music when studying improves learning. They gave two groups of students an identical 2- hour lesson, and then time to study for a short exam. One group of 50 students who told the researchers that they liked classical music, listened to classical music while they studied. The other group of 75 student who told the researchers that they did not like classical music, studied in silence. The results showed that that the students who listened to classical music performed better on the test.
4. A chiropractor wants to know if her adjustments relieve back pain. She performs adjustments on 25 patients with back pain. Afterward, 18 of the patients say they feel better. She concludes that the adjustments are an effective treatment.

Some Additional Comments

- When an observed difference in treatment effects is too large to reasonably have occurred purely by chance, we say that the difference is **statistically significant**. Statistical significance is established using formal inferential statistical methods (confidence intervals and/or tests of significance) which we will see later in the course.
- If significant differences among treatments are found after running a properly designed experiment (that strictly applies these principles, we conclude that the differences are *caused* by the treatments.
- A properly designed experiment minimizes the possibility that the study suffers from confounding.
- A potential weakness of some experiments is the **lack of realism**. This occurs when the experiment results do not reflect what happens in situations wider in scope than the experiment. For example, can results on white mice be generalized to humans?

Discussion Comments

- **One-track experiment:** The experimenter cannot determine if the subjects may be feeling better because of a placebo effect rather than any real effect of 500mg of Nograine. There is no 'baseline' group to compare the Nograine subjects to.
- **ADD:** The experimenters assess politeness in interviews, but because they know which children received the real drug, they may unintentionally speak differently to these children during the interviews. Or, they might interpret the children's behavior differently because they know which subjects received the real drug. These are experimenter effects that can confound the study results. The experiment should have been double-blind.
- **Exercise:** The results of this study will be difficult to interpret because there are so few replications in each group. With a rare disease, however, it is usually difficult to find people to participate in an experiment.
- **Classical Music:** The problem with this study is that students were not randomly assigned to the two groups. By placing students who liked classical music in one group and students who do not like it in the other, the researchers created a situation in which the two groups do not share the same general characteristics. It would have been better to split the two groups into two subgroups with one subgroup exposed to the classical music and the other silence.
- **Chiropractor:** The 25 patients who receive adjustments represent a treatment group, but this study lacks a control group. It is a one-track experiment. The patients may be feeling better because of a placebo effect rather than any real effect of the adjustments. The chiropractor might have improved his study by simulating an adjustment (having no known therapeutic value) with a random subset of the patients, and then compare the two groups to whether or not a placebo effect was involved.