Lesson 22: Is it ever possible to get convincing evidence of a cause-and-effect relationship?

Figure

"Polio pioneers"—some of the many children who took part in trials of poliomyelitis vaccine

Let's read about an experiment.

"A calculated risk": the Salk polio vaccine field trials of 1954

Marcia Meldrum

An Experiment:
Manipulates the factor levels to create treatments.
Randomly assigns subjects to these treatment levels.
Compares the responses of the subject groups across treatment levels.

The Four Principles of Experimental Design

1. Control : we control sources of variation other than the factors that we are testing by making conditions as similar as possible for all treatment groups.
2. Randomize : allows us to equalize the effects of unknown or uncontrollable sources of variation. It does not eliminate the effects of these sources, but it should spread them out across treatment levels so we can see past them.
3. Replicate : replicate experiment over as many subjects as possible. If, as it often happens, the subjects of the experiment are not a representative sample from the population of interest, replicate the entire study with a different group of subjects, preferably from a different part of the population.
4. Block : sometimes, attributes of the experimental units that we are studying and that we can't control may nevertheless affect the outcomes of an experiment. If we group similar individuals together and then randomize within each of these blocks, we can remove much of the variability due to difference among the blocks.

Subjects or participants: humans who are experimented on

Experimental units: other individuals (rats, petri dishes of bacteria etc.)

Levels of a factor: specific values that the experimenter chooses for the factor (explanatory variable)

Treatment: the combination of specific levels from all the factors that the experimental unit receives

Control Treatments vs. Control Group

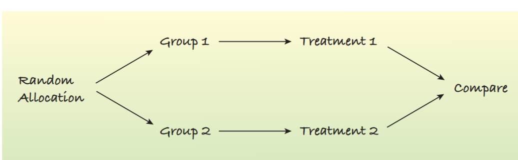
is a baseline measurement

Control treatment is a level of a factor.
We use this to compare the treatment results to a situation in which "nothing happens"

Example: Software that shortens download time.
control treatment: download time without software

the experimental units to whom the control treatment is applied.

A Diagram helps in thinking about the experiment as well as in explaining to others.



In 2007, deaths of a large number of pet dogs and cats were ultimately traced to contamination of some brands of pet food. The manufacturer now claims that the food is safe, but before it can be released, it must be tested.

In an experiment to test whether the food is now safe for dogs to eat, what would be the treatments and what would be the response variable? How will you implement the principles of control, randomization, and replication?

Plan:
Response variable:
Subjects:
Factor(s):
Treatment(s):

Experimental Design

Control:
Replicate:
Random assignment:
Diagram:
Results:

An ad for OptiGro plant fertilizer claims that with this product you will grow “juicier, tastier” tomatoes.You’d like to test this claim, and wonder whether you might be able to get by with half the specified dose. How can you set up an experiment to check out the claim?

Of course, you’ll have to get some tomatoes, try growing some plants with the product and some without, and see what happens. But you’ll need a clearer plan than that. How should you design your experiment?

Plan State what you want to know.

I want to know whether tomato plants grown with OptiGro yield juicier, tastier tomatoes than plants raised in otherwise similar circumstances but without the fertilizer.

Response Specify the response variable.

I'll evaluate the juiciness and taste of the tomatoes by asking a panel of judges to rate them on a scale from 1 to 7 in juiciness and in taste.

Treatments Specify the factor levels and the treatments.

The factor is fertilizer, specifically OptiGro fertilizer. I'll grow tomatoes at three different factor levels: some with no fertilizer, some with half the specified amount of OptiGro, and some with the full dose of OptiGro. These are the three treatments.

Experimental Units Specify the experimental units.

I'll obtain 24 tomato plants of the same variety from a local garden store.

Experimental Design Observe the principles of design:

Control any sources of variability you know of and can control.

I'll locate the farm plots near each other so that the plants get similar amounts of sun and rain and experience similar temperatures. I will weed the plots equally and otherwise treat the plants alike.

Replicate results by placing more than one plant in each treatment group.

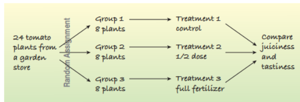
I'll use 8 plants in each treatment group.

Randomly assign experimental units to treatments, to equalize the effects of unknown or uncontrollable sources of variation.

Describe how the randomization will be accomplished.

To randomly divide the plants into three groups, first I'll label the plants with numbers 00–23. I'll look at pairs of digits across a random number table. The first 8 plants identified (ignoring numbers 24–99 and any repeats) will go in Group 1, the next 8 in Group 2, and the remaining plants in Group 3.

Make a Picture A diagram of your design can help you think about it clearly.



Specify any other experiment details. You must give enough details so that another experimenter could exactly replicate your experiment. It's generally better to include details that might seem irrelevant than to leave out matters that could turn out to make a difference.

I will grow the plants until the tomatoes are mature, as judged by reaching a standard color.

I'll harvest the tomatoes when ripe and store them for evaluation.

Specify how to measure the response.

I'll set up a numerical scale of juiciness and one of tastiness for the taste testers. Several people will taste slices of tomato and rate them.

Once you collect the data, you'll need to display them and compare the results for the three treatment groups.

I will display the results with side-by-side boxplots to compare the three treatment groups.

I will compare the means of the groups.

To answer the initial question, we ask whether the differences we observe in the means of the three groups are meaningful.

Because this is a randomized experiment, we can attribute significant differences to the treatments. To do this properly, we'll need methods from what is called "statistical inference," the subject of the rest of this book.

If the differences in taste and juiciness among the groups are greater than I would expect by knowing the usual variation among tomatoes, I may be able to conclude that these differences can be attributed to treatment with the fertilizer.

Blinding

There are two main classes of individuals who can affect the outcome of the experiment:

- those who could influence the results (the subjects, treatment administrators, or technicians)
- those who evaluate the results (judges, treating physicians, etc.)

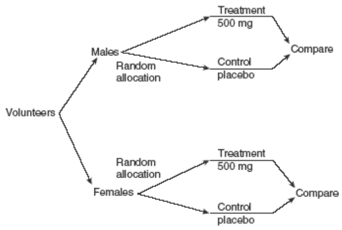
When all the individuals in either one of these classes are blinded, an experiment is said to be **single-blind.** When everyone in *both* classes is blinded, we call the experiment **double-blind.**

Placebos

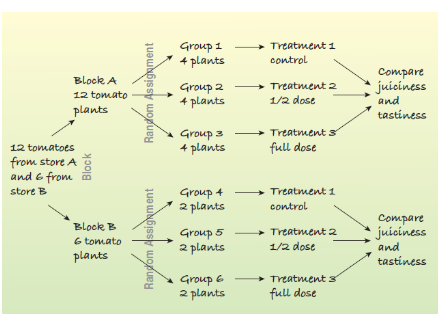
A treatment known to have no effect, administered so that all groups experience the same conditions. Many subjects respond to such a treatment (a response known as a placebo effect). Only by comparing with a placebo can we be sure that the observed effect of a treatment is not due simply to the placebo effect.

Blocking

- Blocking is the same idea for experiments as stratifying is for sampling.
- Both methods group together subjects that are similar and randomize within those groups as a way to remove unwanted variation.
- We use blocks to reduce variability so we can see the effects of the factors; we're not usually interested in studying the effects of the blocks themselves.



Multiple factors in the same experiment can help examine what happens when the factor levels are applied in different *combinations*.



Matching

In a retrospective or prospective study, subjects who are similar in ways not under study may be matched and then compared with each other on the variables of interest. Matching, like blocking, reduces unwanted variation.

For example, a retrospective study of music education and grades might match each student who studies an instrument with someone of the same sex who is similar in family income but didn't study an instrument. When we compare grades of music students with those of non-music students, the matching would reduce the variation due to income and sex differences.

Matched Pairs Design (stat-trek website)

A matched pairs design is a special case of a randomized block design. It can be used when the experiment has only two treatment conditions; and subjects can be grouped into pairs, based on some blocking variable. Then, within each pair, subjects are randomly assigned to different treatments.

The table belows shows a matched pairs design for a hypothetical medical experiment, in which 1000 subjects each receive one of two treatments - a placebo or a cold vaccine. The 1000 subjects are grouped into 500 matched pairs. Each pair is matched on gender and age. For example, Pair 1 might be two women, both age 21. Pair 2 might be two men, both age 21. Pair 3 might be two women, both age 22; and so on.

	Treatment	
Pair	Placebo	Vaccine
1	1	1
2	1	1
...	...	...
499	1	1
500	1	1

Experiments vs. Sample Surveys

Both experiments and sample surveys use randomization to get unbiased data.

Experiments try to assess the effects of treatments. Experimental units are not always drawn randomly from the population.

Sample surveys try to estimate population parameters, so the sample needs to be as representative of the population as possible

We want a sample to exhibit the diversity and variability of the population, but for an experiment the more homogeneous the subjects the more easily we'll spot differences in the effects of the treatments

Unless the experimental units are chosen from the population at random, you should be cautious about generalizing experiment results to larger populations until the experiment has been repeated under different circumstances

Experiments deal better with biases.

If all groups have biases and they're applied treatment to ...biases cancel out.

Confounding

When the levels of one factor are associated with the levels of another factor, we say that these two factors are **confounded.**

Example:

Professor Stephen Ceci of Cornell University performed an experiment to investigate the effect of a teacher's classroom style on student evaluations. He taught a class in developmental psychology during two successive terms to a total of 472 students in two very similar classes. He kept everything about his teaching identical (same text, same syllabus, same office hours, etc.) and modified only his style in class. During the fall term, he maintained a subdued demeanor. During the spring term, he used expansive gestures and lectured with more enthusiasm, varying his vocal pitch and using more hand gestures. He administered a standard student evaluation form at the end of each term

The students in the fall term class rated him only an average teacher. Those in the spring term class rated him an excellent teacher, praising his knowledge and accessibility, and even the quality of the textbook. On the question "How much did you learn in the course?" the average response changed from 2.93 to 4.05 on a 5-point scale.

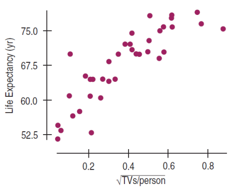
How much of the difference he observed was due to his difference in manner, and how much might have been due to the season of the year? Fall term in Ithaca, NY (home of Cornell University), starts out colorful and pleasantly warm but ends cold and bleak. Spring term starts out bitter and snowy and ends with blooming flowers and singing birds. Might students' overall happiness have been affected by the season and reflected in their evaluations

Lurking or Confounding

A lurking variable is usually thought of as a variable associated with both y and x that makes it appear that x may be causing y.

A confounding variable is associated in a noncausal way with a factor and affects the response. Because of the confounding, we find that we can't tell whether any effect we see was caused by our factor or by the confounding variable—or even by both working together.

Example



Lurking: Higher Living standard

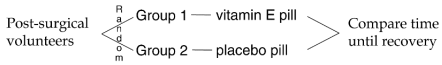
Example: Professor Ceci's class.

confounding:

is the response because of weather of his teaching or both

Healing. A medical researcher suspects that giving post-surgical patients large doses of vitamin E will speed their recovery times by helping their incisions heal more quickly. Design an experiment to test this conjecture. Be sure to identify the factors, levels, treatments, response variable, and the role of randomization.

Healing.



Answers will vary. This double-blind experiment has 1 factor (vitamin E), at 2 levels (vitamin E and no vitamin E), resulting in 2 treatments. The response variable measured is the time it takes the patient to recover from the surgery. Randomly select half of the patients who agree to the study to get large doses of vitamin E after surgery. Give the other patients in the study a similar looking placebo pill. Monitor their progress, recording the time until they have reached an easily agreed upon level of healing. Have the evaluating doctor blinded to whether the patient received the vitamin E or the placebo. Compare the number of days until recovery of the two groups.

SAT Prep. Can special study courses actually help raise SAT scores? One organization says that the 30 students they tutored achieved an average gain of 60 points when they retook the test.

a) Explain why this does not necessarily prove that the special course caused the scores to go up.
b) Propose a design for an experiment that could test the effectiveness of the tutorial course.
c) Suppose you suspect that the tutorial course might be more helpful for students whose initial scores were particularly low. How would this affect your proposed design?