

Experiments

In an *experiment*, a researcher assigns *treatments* to a group of *experimental units*. In a typical clinical trial or similar study, the experimental units are people who volunteer for or are recruited into the study. However, the same terminology is used if the experimental units are not people, e.g. they may be animals, plants, pieces of industrial apparatus, or just about anything else that could be involved in an experiment.

A key point in most experiments is a *control group*, or a group of experimental units selected under similar conditions to the group receiving the treatment of interest.

Often the control group receives a *placebo*.

The placebo effect —

Sometimes people with a disease are cured *even though* they only received the placebo, maybe because of the psychological effect of receiving a drug that they believe will cure them. This shows one reason why it is important to have a control group — otherwise the people given the new drug may appear cured because of the placebo effect, when in reality it has no effect.

Historical controls are also used in some studies but this is also a bad idea.

Randomization is important for several reasons:

- To eliminate the bias that might occur if the researchers simply chose who would receive what treatment
- To balance the groups on variables that might affect the response
- To eliminate lurking variables

It is almost always a good idea to *blind* the study, i.e. subjects do not know which treatment they are receiving. Sometimes trials are *double-blinded*.

The result of an experiment is said to be *statistically significant* if the difference between the treatments is large enough that it cannot be attributed to chance. Usually, in clinical trials it is necessary to establish a statistically significant result before a new drug can be licensed. This affects the *sample size* needed to conduct the experiment.

Replication refers to repeating the same experiment on many experimental units so as to increase the chance of gaining a statistically significant result.

Repetition of the whole experiment (e.g. at a different institution) is usually needed before the results are generally accepted by the scientific community.

Other types of experimental design

Multi-factor experiments:

Refers to experiment when there is more than one type of treatment.

Example: In a study of patients who are at high risk of heart disease, we might want to compare two different types of treatment: low fat diet *or* cholesterol-reducing drugs. We might want to decide whether either or both of these is effective. So we could set up an experiment in which we divide patients into *four* groups:

- Group A: Low-fat diet, no cholesterol mediation (or a placebo)
- Group B: Regular diet, no cholesterol mediation (or a placebo)
- Group C: Low-fat diet plus cholesterol mediation
- Group D: Regular diet plus cholesterol mediation

As in more simple experiments, we would use randomization to decide which patient goes in which group.

We could add more factors, for example a third factor could be exercise (half the patients in each group put on a high-exercise regime, the other either told not to exercise or not given any advice about the subject). The disadvantage tends to be that as we add more and more factors, it requires more patients to get an adequate number in each group.

Matched pairs design:

Suppose we are interested in whether taking a sports drink before practice for an athletic event improves performance.

A possible experiment is as follows. At one training session, randomly select half the athletes to receive the drink, while the others get nothing. Then measure their performance (e.g. if it's football players, the “performance” might be their time in a 40 yard dash)

Next practice session, reverse the groups who do or do not get the sports drink and repeat the trial.

If when the two trials are combined, the athletes on average did better when they received the sports drink than when they did not, you can conclude that the drink had an effect.

This is a *matched pairs* design because each sampling unit consisted of a pair of observations: one performance by an athlete using the drink, and another by *the same* athlete without the drink. The fact that the experiment uses the same athlete twice avoids biases caused by the fact that the athletes do not all have the same ability.

It is also called a *crossover design*.

Suppose we only had one session to perform the entire experiment?

Blocking in an experiment

In medical studies, often the patients are not all alike before they enter the experiment — e.g. different age groups, different medical conditions, mixture of men and women (if gender is relevant). If we just threw them all randomly into the study, the results might be biased because e.g. by chance, the patients receiving the placebo tend to be older than those receiving the new drug.

We could reduce this bias by *blocking*, i.e. first group the patients into “blocks” matched by age and medical condition, then within each block, assign the treatments and controls at random.

The resulting experiment is called a *randomized blocks* experiment.

If the blocks are of size two, this takes us back to a matched pairs design.

Retrospective and Prospective Studies

A study is *prospective* if all the procedures for the study (including the selection of subjects) are laid out at the beginning of the study. It is *retrospective* if the subjects are only recruited *after* the events of interest have taken place.

Experiments are, by their nature, prospective.

However observational studies could be prospective or retrospective. The Six Cities Study was an example of a prospective study because, although the study itself lasted 20 years, the participants were all selected in one group at the beginning and the detailed plan of the study was laid down by the researchers at that time.

Usually prospective studies are more useful than retrospective studies because they are better controlled. However, there is one exception to this rule, which is a *case-control study*.

Case-control Studies

Example: In 1950, Doll and Hill in England examined 709 patients who had lung cancer. At the same time they recruited 709 patients who did not have cancer but who were matched with the 709 patients who did, in the sense that they were in the same hospital, the same gender, and the same 5-year age group. They then found that of the cancer patients, 688 were smokers and 21 were not. Of the non-cancer patients, 650 were smokers and 59 were not. The lung cancer rate was higher among smokers than non-smokers (to a statistically significant extent).

Note that the study tells us *nothing* about the chances that someone who either is or is not a smoker gets cancer. But it is very good at establishing an *association* which, in this case, contributed to the general acceptance of a cause and effect relationship.

Many environmental health studies use a case-control format because they avoid some of the problems of pure-observational studies, and also the ethical and practical issues associated with clinical trials. However they don't avoid the issue of lurking variables (e.g. if there was a tendency for smokers also to be heavy consumers of alcohol, and it was really alcohol that "caused" lung cancer, the study would not be able to detect that).