Experimental design

Note: You can see this plus a bit more online at:

http://oac.med.jhmi.edu/humane_site/topics/12A.html

Basic principles

1. Formulate question/goal in advance
2. Comparison/control
3. Replication
4. Randomization
5. Stratification (aka blocking)
6. Factorial experiments

Example

Question: Does salted drinking water affect blood pressure (BP) in mice?

Experiment:

1. Provide a mouse with water containing 1% NaCl.
2. Wait 14 days.
3. Measure BP.

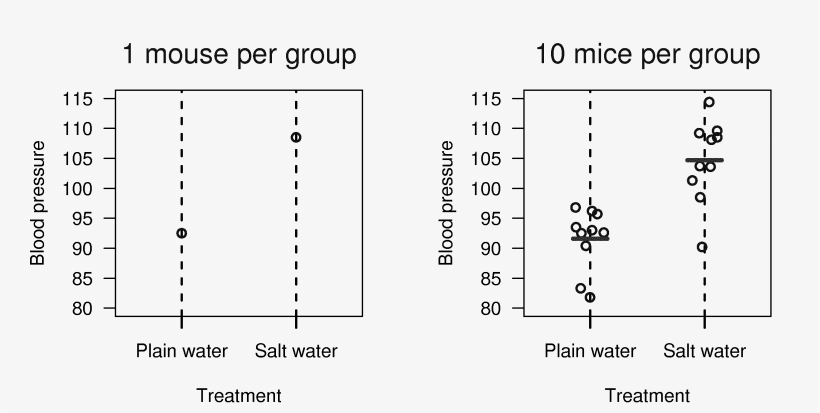
Comparison/control

Good experiments are comparative.

- Compare BP in mice fed salt water to BP in mice fed plain water.
- Compare BP in strain A mice fed salt water to BP in strain B mice fed salt water.

Ideally, the experimental group is compared to concurrent controls (rather than to historical controls).

Replication



Why replicate?

- Reduce the effect of uncontrolled variation (i.e., increase precision).
- Quantify uncertainty.

A related point:

An estimate is of no value without some statement of the uncertainty in the estimate.

Randomization

Experimental subjects ("units") should be assigned to treatment groups at random.

At random does not mean haphazardly.

One needs to explicitly randomize using

- A computer, or
- Coins, dice or cards.

Why randomize?

- Avoid bias.
  - For example: the first six mice you grab may have intrinsically higher BP.
- Control the role of chance.
  - Randomization allows the later use of probability theory, and so gives a solid foundation for statistical analysis.

Stratification

- Suppose that some BP measurements will be made in the morning and some in the afternoon.
- If you anticipate a difference between morning and afternoon measurements:
  - Ensure that within each period, there are equal numbers of subjects in each treatment group.
  - Take account of the difference between periods in your analysis.
- This is sometimes called “blocking”.

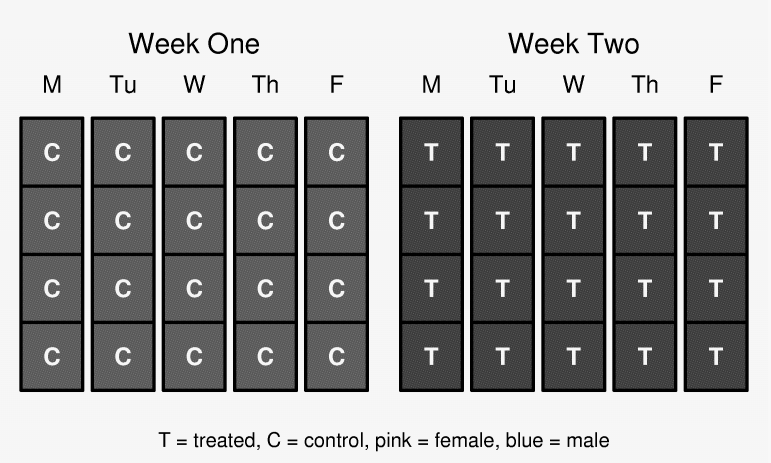
Example

- 20 male mice and 20 female mice.
- Half to be treated; the other half left untreated.
- Can only work with 4 mice per day.

Question: How to assign individuals to treatment groups and to days?

Johns Hopkins University Center for Alternatives to Animal Testing

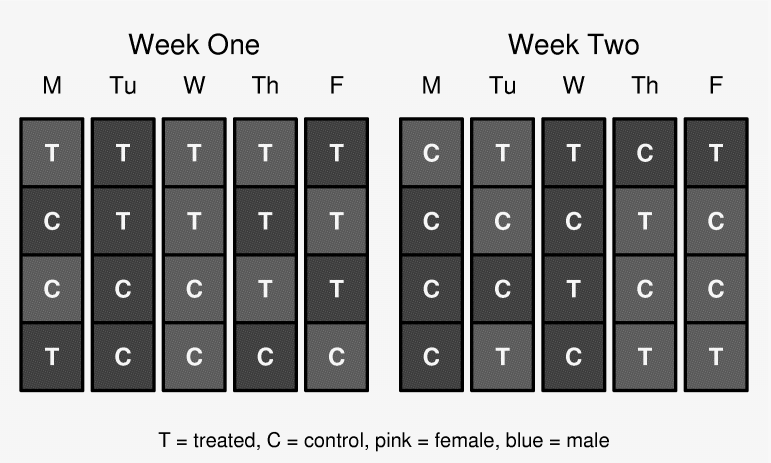
An extremely bad design



T = treated, C = control, pink = female, blue = male

Johns Hopkins University Center for Alternatives to Animal Testing

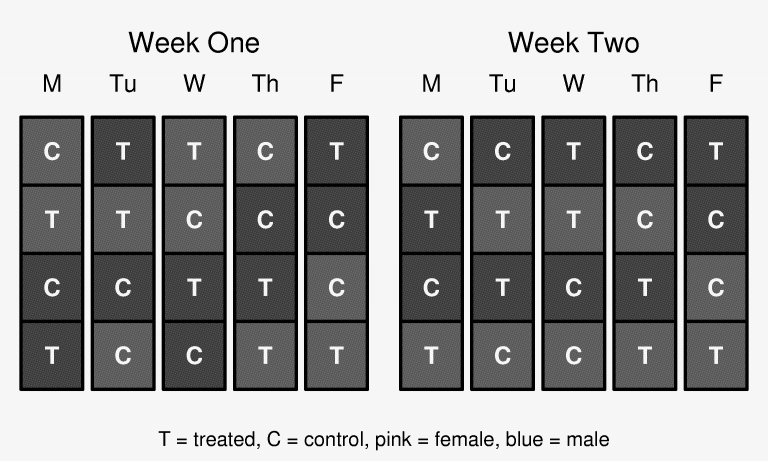
Randomized



T = treated, C = control, pink = female, blue = male

ENHANCING HUMANE SCIENCE IMPROVING ANIMAL RESEARCH

A stratified design



T = treated, C = control, pink = female, blue = male

Randomization and stratification

- If you can (and want to), fix a variable.
  - e.g., use only 8 week old male mice from a single strain.
- If you don't fix a variable, stratify it.
  - e.g., use both 8 week and 12 week old male mice, and stratify with respect to age.
- If you can neither fix nor stratify a variable, randomize it.

Factorial experiments

Suppose we are interested in the effect of both salt water and a high-fat diet on blood pressure.

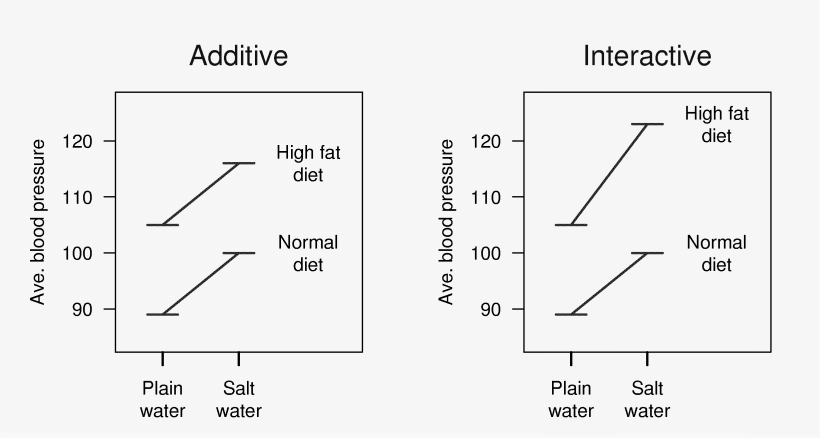
Ideally: look at all 4 treatments in one experiment.

Plain water	□	Normal diet
Salt water		High-fat diet

Why?

- We can learn more.
- More efficient than doing all single-factor experiments.

Interactions



Other points

- Blinding
  - Measurements made by people can be influenced by unconscious biases.
  - Ideally, dissections and measurements should be made without knowledge of the treatment applied.
- Internal controls
  - It can be useful to use the subjects themselves as their own controls (e.g., consider the response after vs. before treatment).
  - Why? Increased precision.

Other points

- Representativeness
  - Are the subjects/tissues you are studying really representative of the population you want to study?
  - Ideally, your study material is a random sample from the population of interest.

Summary

Characteristics of good experiments:

- Unbiased
  - Randomization
  - Blinding
- High precision
  - Uniform material
  - Replication
  - Stratification
- Simple
  - Protect against mistakes
- Wide range of applicability
  - Deliberate variation
  - Factorial designs
- Able to estimate uncertainty
  - Replication
  - Randomization

Salk vaccine trial

1916: first polio epidemic in the US

next 40 years: hundreds of thousands of victims

By 1950s: several vaccines developed; that by Jonas Salk appears most promising

1954: Public Health Service and Nat'l Fdn for Infantile Paralysis (NFIP) ready to test the Salk vaccine in a field trial

See Freedman, Psiani, Purves (1998) *Statistics*, 3rd ed, Ch 1–2

Possible designs for the vaccine trial

1. Give the vaccine to many children and look at the rate vs the previous year.
2. Compare those vaccinated to those whose parents refused vaccination.
3. Vaccinate grade 2 (in consenting) and compare to grades 1 and 3. [This is what the NFIP chose to do.]
4. Vaccinate some portion (chosen at random) of those whose parents consent.

Best study:

double-blind randomized placebo-controlled

Results of 1954 Salk vaccine trial

The randomized controlled double-blind experiment

	Size	Rate
Treatment	200,000	28
Control	200,000	71
No consent	350,000	46

The NFIP study

	Size	Rate
Grade 2 (vaccine)	225,000	25
Grades 1 & 3 (control)	725,000	54
Grade 2 (no consent)	125,000	44

Note: Rates are per 100,000

Points

- NFIP study: vaccine appears to lower rate 54 $\rightarrow$ 25 (vs 71 $\rightarrow$ 28).

  The control group included children whose parents would not have consented.

- Might the vaccine have no effect? (Could the observed differences be simply chance variation?)
  - In the randomized controlled trial, it is relatively simple to answer this question, as the role of chance was according to our design.
  - In the NFIP study, it is impossible to tell, as chance is not under our control.

The portacaval shunt

A long, hazardous surgery to treat cirrhosis of the liver.

Do the benefits outweigh the risks?
Over 50 studies have considered this.

	Degree of enthusiasm		
Design	Marked	Moderate	None
No controls	24	7	1
Controls, but not randomized	10	3	2
Randomized controlled	0	1	3

In the studies where the controls were not chosen at random, sicker patients were chosen as controls.

Historical controls

Historical controls: patients treated the old way in the past.

Problem: treatment group and historical control group may differ in important ways besides the treatment.

	Randomized controlled		Historically controlled	
	+	–	+	–
Coronary bypass surgery	1	7	16	5
5-FU	1	7	2	0
BCG	2	2	4	0
DES	0	3	5	0