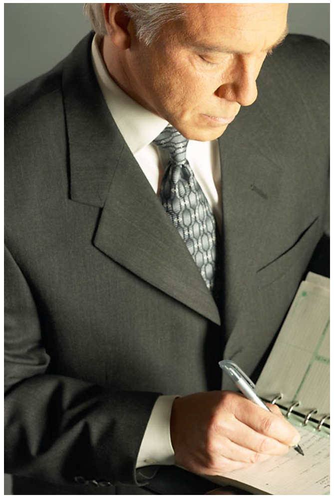
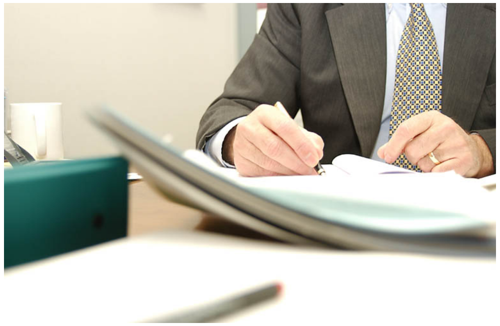
1.3 Types of studies:

Observational Studies
and
Experiments

Observational Studies

- Simply observing what happens
  - An opinion sample survey is an observational study.
  - Researcher does not 'impose' a treatment or randomly assign subjects to a group.





Study: Tanning and Skin Cancer

- The **observational study** involved 1,500 people.
- Selected a group of people who had skin cancer and another group of people who did not have skin cancer
  - Group membership based on innate characteristics in the subjects (they were not assigned to a group).
- Asked all participants whether they used tanning beds.
  - Wanted to see if there was an association between tanning beds and skin cancer prevalence.

Study: Sodium and Blood Pressure

- Enroll 100 individuals in the **observational study.**

- Give them diet diaries where they record everything they eat each day. From this the amount of sodium in the diet is found.

- Not assigning subjects to any certain amount of sodium (they choose it on their own).

- Measure their blood pressure.

Types of observational studies

- **Retrospective** – look at past records and historical data.
  - Tanning and Skin Cancer
  - Can be a *Case-Control study*
- **Prospective** – identify subjects and collect data as events unfold.
  - Sodium and Blood Pressure
  - If you collect data at multiple time points, it is a *longitudinal study*.

- Observational studies are often used in marketing or health studies.

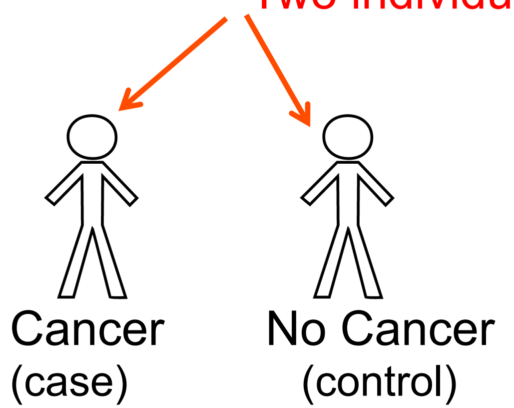
Case-Control study

- The conclusions for **observational studies** are not as strong as experiments (which use randomization), but sometimes we can not ethically perform an experiment.
  - Can you randomly assign individuals to be in a cigarette smoking group? No.
  - Can you randomly assign someone to use a tanning bad if there is a risk of cancer? No.
- Matched Case-Control studies can help give credibility to relationships in such situations.

Matched Case-Control study

- Tanning beds and skin cancer
  - Cases (cancer) and controls (no cancer)
  - Match individuals on relevant characteristics…
    - Same age
    - Same outdoor tanning behaviors
    - Similar diet

Two individuals similar in many ways.



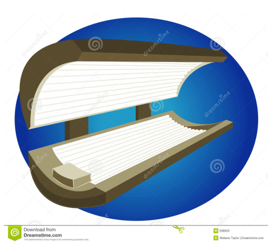
Matched Case-Control study

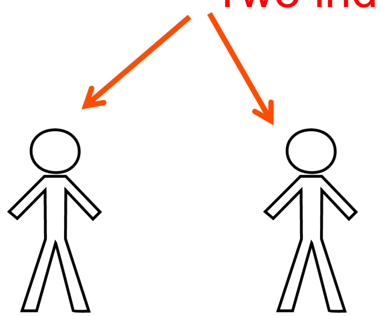
- We try to ‘match’ cases and controls who have similar characteristics to overcome the study limitations due to **confounding** (next topic).
  - Then we ask about their tanning bed use.
- If *everything else* seems to be the same between these two people EXCEPT tanning bed use, and the tanner has cancer, then that gives some credibility to cancer and tanning beds be related.

Matched Case-Control study

- It gives a rebuttle to the devil's advocate saying the cancer could have been due to age, it could have been diet, ...
  - The cancer is *not* due to age (they're the same)
  - The cancer is *not* due to diet (they're the same)

Two individuals similar in many ways.





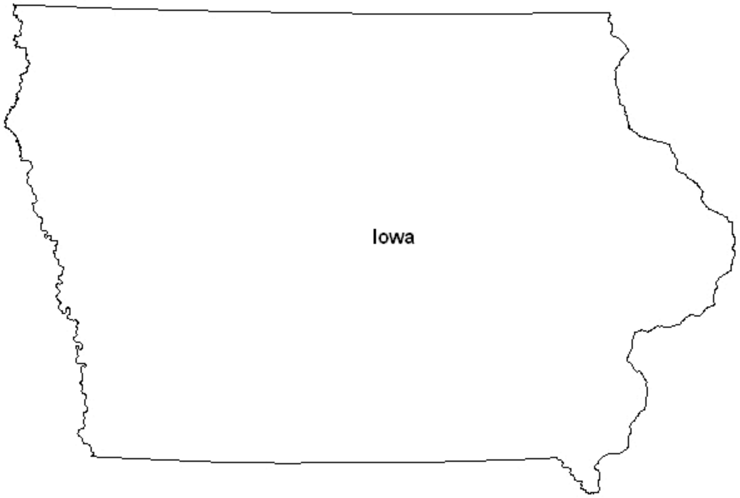
Cancer (case)

No Cancer (control)

Matched Case-Control study

- Using a matched case-control study is not as strong as being able to use randomization, as in an experiment, but if it's not ethical to assign treatment groups, it can increase the strength of the conclusions of the study.

Confounding



- Fertilizer and Corn yield
  - All plots on the west side of state are in fertilizer A.
  - All plots in the middle of the state are in fertilizer B.
  - All plots on the east side of state are in fertilizer C.

PROBLEM: If we see differences in yield between the fertilizer groups, is it due to the **fertilizer** or is it due to the **soil**? **We don't know.**

This is because 'Fertilizer' and 'Soil' are **confounded** in the scenario (bad thing here).

Confounding

- Sodium and blood pressure.
  - All subjects on the low sodium diet had their blood pressure measured by a registered nurse using a standard manual cuff and stethoscope.
  - All subjects on the high sodium diet had their blood pressure measured using an automated cuff and digital readout.

PROBLEM: If we see differences in blood pressure between the groups, is it due to the **diet** or is it due to the **method of measurement**? **We don't know**

This is because 'Diet' and 'Method of Measurement' are **confounded** in the scenario (bad thing here).

Confounding

- A problem in observational studies.

- Something we try to overcome in observational studies by using ‘matched pairs’ of case-control, but almost always still part of the argument.

- Experimental studies (next topic) uses randomization as a tool to fight the occurrence of confounding.

Experiments

- Explanatory variable – Factor.
  - e.g. Dosage (0ml, 10ml, 20ml)

    There are 3 **treatments** in this experiment.

- Response variable.
  - e.g. Percent of skin irritation alleviated.

- Randomly assign subjects to treatment groups.

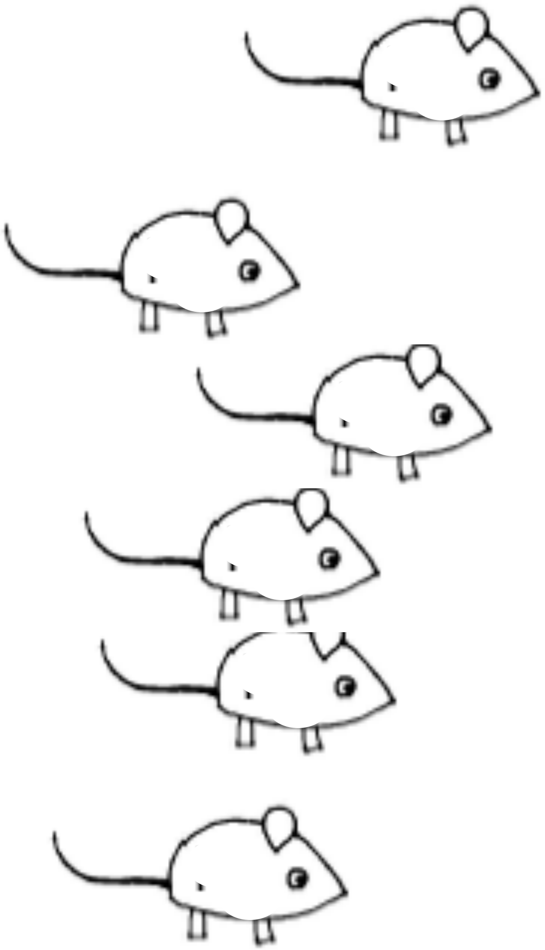
Experiments

- The experimenter must *actively* and *deliberately* manipulate the factor(s) to establish the method of treatment.

- **Experimental units are assigned at random to the treatments.**

Experiments

n = 6 mice to be randomly assigned to 2 groups



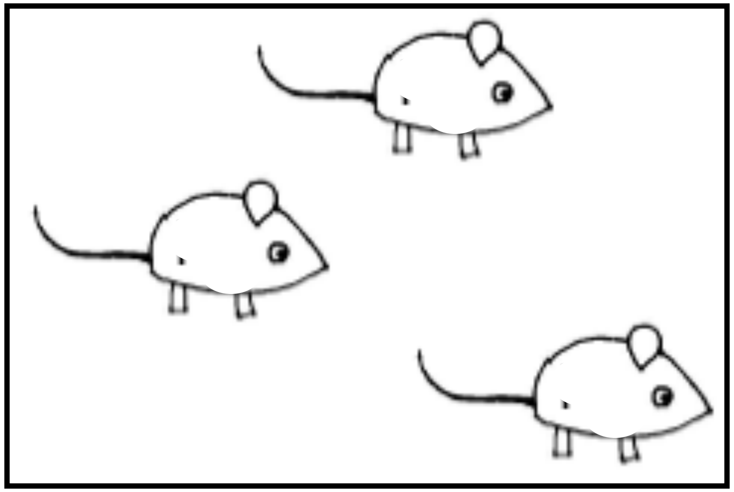
Present Drug Group

New Drug Group

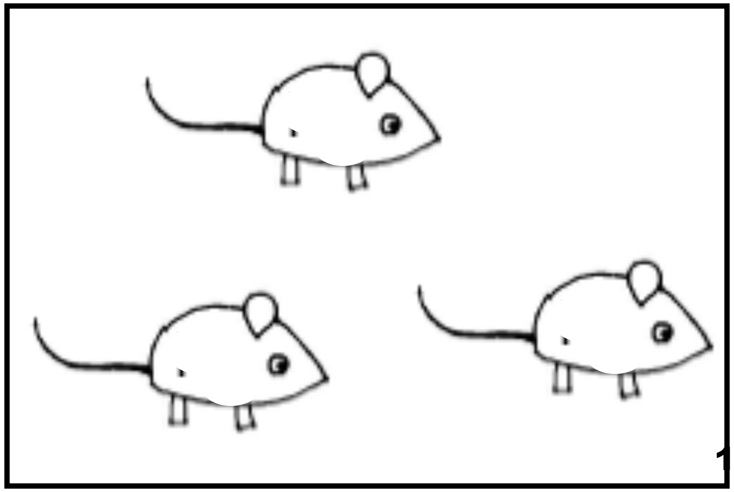
Experiments

n = 6 mice to be randomly assigned to 2 groups

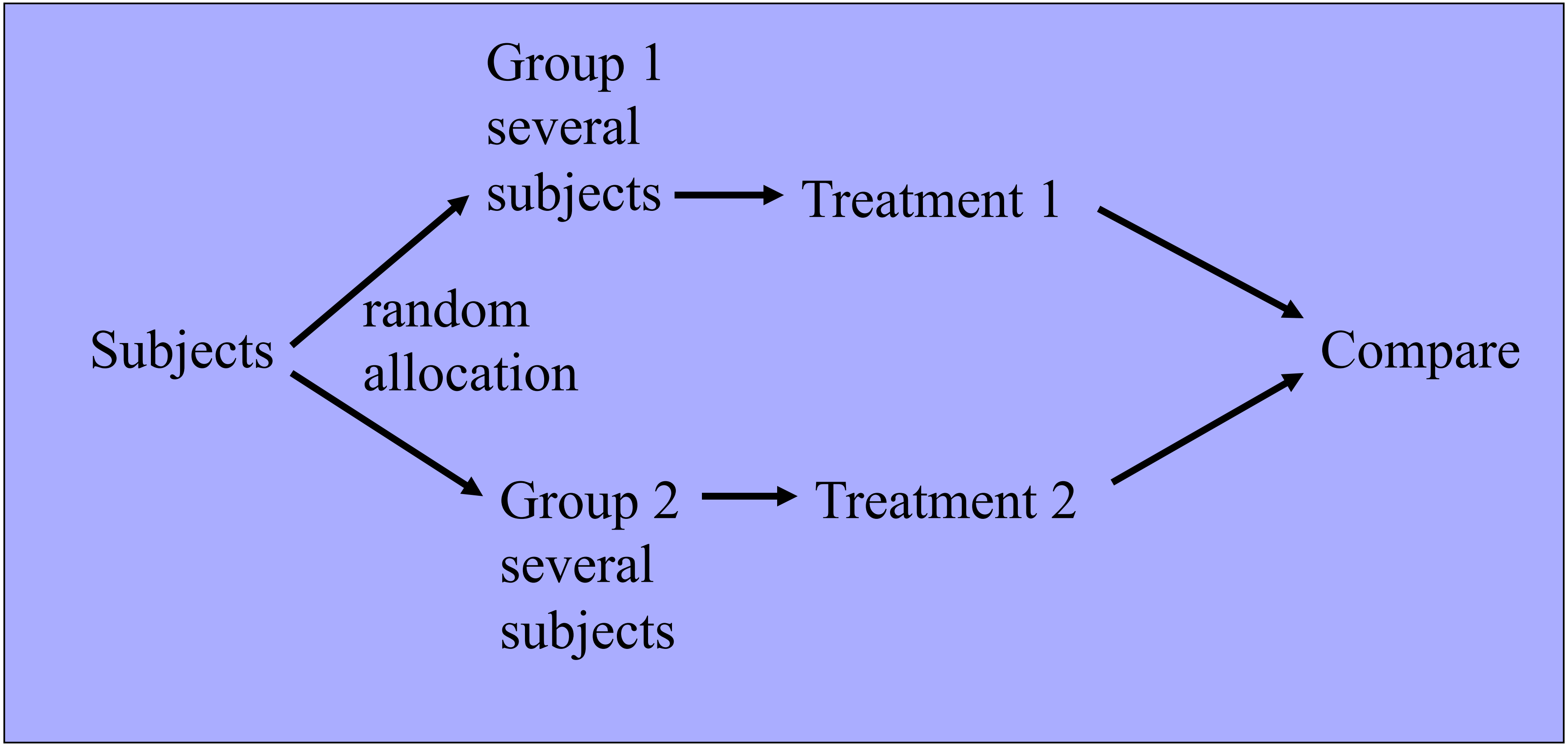
Present Drug Group



New Drug Group



Diagrams


Group 1 several subjects
Treatment 1
Subjects
random allocation
Compare
Group 2 several subjects
Treatment 2

Experiment:
Sodium and Blood Pressure

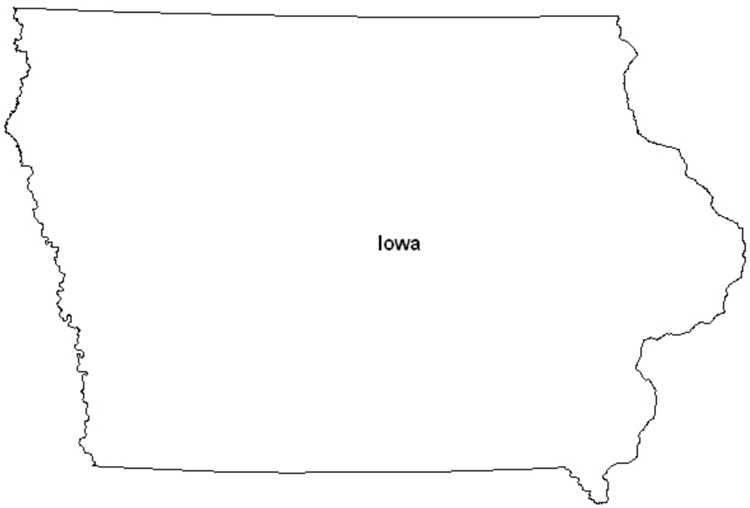
- 20 subjects.
- Factor – amount of dietary sodium.
- Treatments: low sodium diet and high sodium diet.
- 10 subjects randomly assigned to each treatment.
- Response – systolic blood pressure.

Experiment: Sodium and Blood Pressure

- One factor: Amount of dietary sodium
- Two treatments:
  - 1) High sodium diet
  - 2) Low sodium diet

Why does randomization help?

- The effects of other factors (besides the variable of interest) on the response gets ‘averaged-out’ across the treatment groups.



- All fertilizers get some good soil and some bad soil.
  - **Fertilizer** is no longer confounded with **soil**.

Randomization

- Randomization tends to spread the effects of uncontrolled outside variables evenly across the treatment groups.
  - e.g. Make sure sick & healthy people are in *both* treatment groups of a drug experiment.
  - e.g. Make sure young and old people are in both treatment groups of a drug experiment.

- We can’t completely know which things will affect the response, so we randomize to try to cover ALL the possibilities.

Randomization

- Randomization reduces the chance that an uncontrolled outside variable will bias the results.

- If units are not assigned at random to treatments, the researcher *can not* use the powerful methods of statistics to draw conclusions.

Inclusion of a Control Group

- A Control Group represents a treatment group that is 'given nothing'.
  - e.g. Dosage (**0ml**, 10ml, 20ml) then the 0ml group is the control group.
  - It's the treatment where 'nothing happens'.
  - We want to give the control group a **placebo** to mimic whatever is happening in the non-control groups so everyone is treated similarly.
    - e.g. shots… give a salt water injection.
    - e.g. acupuncture… insert needles at the 'non-acupuncture' locations.

Inclusion of a Control Group

- We use a control group to show it is the variable of interest (not other confounding variables) causing the change in the response.
  - e.g. 1) Depression and depression drug
    - Treatments: 10ml of drug, placebo
  - e.g. 2) Depression and depression drug
    - Treatments: 10ml of drug, placebo, no drug
    - Can compare more effects

Placebo effect

- Occurs when people believe they are receiving a useful and beneficial treatment.
- Single-blind study: subject does not know which group to which they belong.
- Double-blind study: neither subject nor experimenter knows the group assignments.

Some other terminology

- Subject
  - People, animals, or objects chosen for sample
- Explanatory Variable
  - May explain or cause an effect, is a predictor
- Response Variable
  - Responds to changes in the explanatory variable, the thing we want to predict

Some other terminology

- Variable
  - Item that can take on many values.
    - Age, weight, treatment group.
- Variable(s) of interest
  - Quantities that are the focus of the study.
- Meta-analysis
  - Investigating <u>many</u> past studies to form conclusions about variables of interest.
  - Pools information across studies.