

2. Collecting data

Two fundamentally different ways of collecting data:

Observational study: Passive - record data without interfering with the process or system being observed.

Example: Gallup poll

Experiment: Active - make carefully controlled changes to the system and record response.

Example: Salk vaccine field trial

Observational study and experiment differ fundamentally in the kind and strength of conclusions that can be drawn.

In particular, it is impossible to conclusively establish causality from an observational study - a weakness of epidemiology.

Distinguish between *providing evidence for decision making under uncertainty* and *establishing scientific truth*.

2.2 Experiments

Study the effect of one or more *independent variables* (or *factors*) on a *dependent variable* (or *response*).

Treatment: Combination of specific values of the factors.

The Salk vaccine field trial

Childhood polio claimed hundreds of thousands of victims in the first part of the 20th century.

Jonas Salk developed vaccine in the early fifties.

By 1954, Public Health Service was ready to try vaccine in the real world.

Simple idea: Just vaccinate all the children and see what happens.

Not a good idea! Why?

The Salk vaccine field trial

- Trial was carried out in selected school districts throughout the country where risk of polio was high.
- Subjects: 1.5 million children in grades 1-3 whose parents gave permission for vaccination. Another 0.5 million parents refused.
- The 1.5 million subjects were randomly divided into a treatment group of 0.5 million and a control group of 1 million children (*randomized, controlled experiment*).
- The children in the treatment group received the vaccine. The children in the control group received an injection of saline solution (*Blind experiment*).
- Diagnosticians did not know group membership of subjects (*Double blind experiment*).

Randomized, controlled, double blind experiment

Results:

Polio rates per 100,000 in each group:

Treatment: 28

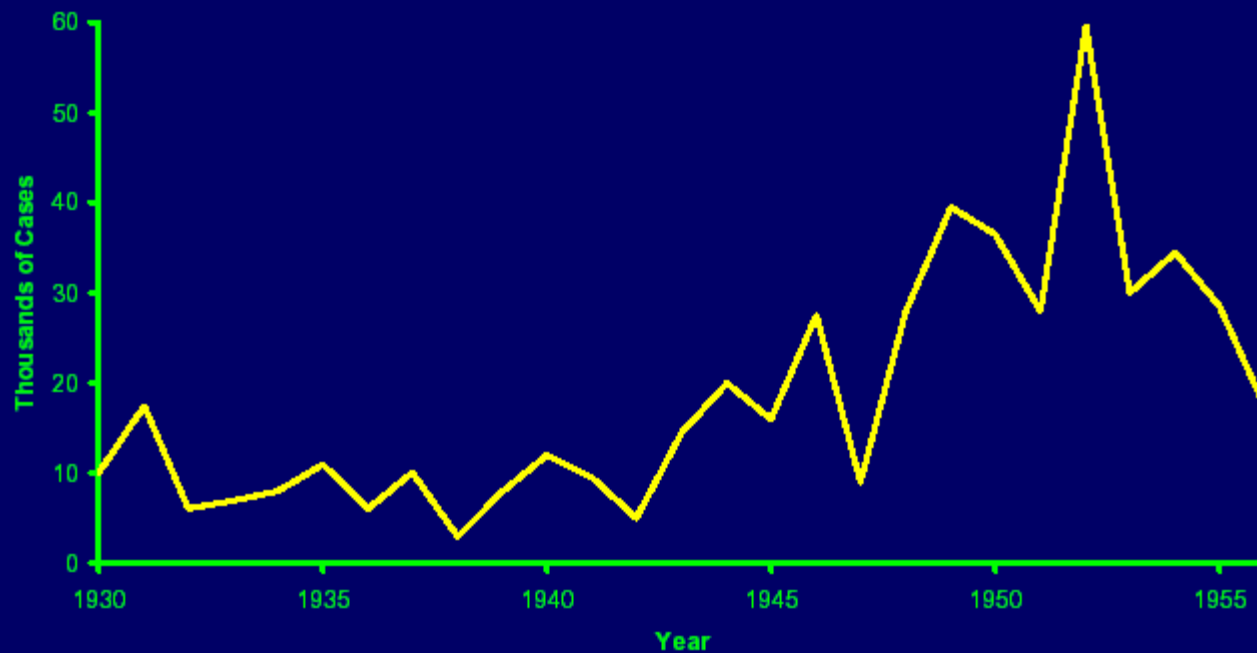
Control: 71

No consent: 46

Questions

1. Why have a control group and not simply use last year's incidence rate for comparison?
2. Why not use children as controls whose parents refuse consent?
3. Why randomize?
4. Why blinding of subjects?
5. Why blinding of diagnosticians?

Number of Polio Cases by Year



Answers

1. Incidence rates varied widely from year to year.
2. Children whose parents refuse vaccination differ systematically from children whose parents allow vaccination. They come from less educated and poorer families where the incidence of polio is *lower*.

Confounding: Treatment and control groups differ with respect to some factor other than treatment.

3. Randomization makes systematic differences between treatment and control groups unlikely - basis of *permutation tests* (later).
4. Subjects were blinded to prevent placebo effect.
5. Diagnosticians were blinded to prevent subconscious bias.

Randomization is key for conclusively establishing a causal relationship

Otherwise, subjects might differ among themselves in crucial ways besides the treatment (confounding).

- Ultrasound and low birth weight.
- Smoking and lung cancer (?)
- Diet and heart attacks.
- Coffee consumption and liver cancer.

Randomization is not a cure-all

- We typically want to generalize experimental results from the subjects to a larger population.
- This is only justified if the subjects are representative for the population, for example if they are a random sample.