

## What is *Statistics*?

*The science of collecting, summarizing and analyzing data.*

*In particular, Statistics is concerned with drawing inferences from a sample about the population from which the sample was drawn.*

(\* ) Collection of data — Experiments and Observational studies.

(\* ) Summarizing data — Descriptive statistics.

(\* ) Analyzing data — Inferential statistics.

## Data Collection: Controlled Experiments

### Salk polio vaccine trials.

(a)

*The NFIP study*

---

	<i>Size</i>	<i>Rate</i>
Grade 2 (vaccine)	225,000	25
Grades 1 and 3 (control)	725,000	54
Grade 2 (no consent)	125,000	44

(\*) Treatment group vs. Control group.

(\*) *Goal* – the treatment and control groups should differ only in the treatment, vaccine vs. no vaccine in this case.

(\*) In a *controlled* experiment, the researchers *control* who goes into each group.

(b)

*The randomized controlled  
double-blind experiment*

---

	<i>Size</i>	<i>Rate</i>
Treatment	200,000	28
Control	200,000	71
No consent	350,000	46

(\*) *Randomized controls* – subjects are assigned to each group randomly.

(\*) *Blind* – subjects do not know which group they are in.

(\*) *Double-blind* – researchers do not know which group each subject is in during the experiment

## Portacaval shunt studies

Table 2. A study of 51 studies on the portacaval shunt. The well-designed studies show the surgery to have little or no value. The poorly-designed studies exaggerate the value of the surgery.

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

Source: N. D. Grace, H. Muench, and T. C. Chalmers, "The present status of shunts for portal hypertension in cirrhosis," *Gastroenterology* vol. 50 (1966) pp. 684–91.

Table 3. Randomized controlled experiments vs. controlled experiments that are not randomized. Three-year survival rates in studies of the portacaval shunt. (Percentages are rounded.)

	<i>Randomized</i>	<i>Not randomized</i>
Surgery	60%	60%
Controls	60%	45%

## Historical controls.

Table 4. A study of studies. Four therapies were evaluated both by randomized controlled trials and by trials using historical controls. Conclusions of trials were summarized as positive (+) about the value of the therapy, or negative (−).

<i>Therapy</i>	<i>Randomized controlled</i>		<i>Historically controlled</i>	
	+	−	+	−
Coronary bypass surgery	1	7	16	5
5-FU	0	5	2	0
BCG	2	2	4	0
DES	0	3	5	0

Note: 5-FU is used in chemotherapy for colon cancer; BCG is used to treat melanoma; DES, to prevent miscarriage.

Source: H. Sacks, T. C. Chalmers, and H. Smith, “Randomized versus historical controls for clinical trials,” *American Journal of Medicine* vol. 72 (1982) pp. 233–40.

(\*) Different experiments conducted to test the same therapy can lead to opposite conclusions depending on the design.

## Data Collection: Observational studies

(\* ) Some studies would be unethical to conduct as controlled experiments. E.g., the effects of smoking or alcohol on a person's health.

(\* ) In a controlled experiment, researchers assign subjects to treatment and control groups.

(\* ) In an observational study, subjects 'select' themselves for treatment or control.

(\* ) Problem: The treatment and control groups may differ in attributes other than the 'treatment' being studied...

## The Clofibrate Trial.

(\*) The efficacy of Clofibrate in preventing heart-attacks was tested in a randomized, double-blind experiment.

(\*) Result: 20% of the subjects in the treatment group and 21% of the subjects in the control group died during the follow-up period.

(\*) Possible explanation for ineffectiveness: Non-adherence of subjects in the treatment group.

(\*) 15% of the adherers in the treatment group died compared to 28% of the non-adherers.

(\*) Observational study! There may be **confounding variables**.

Table 1. The clofibrate trial. Numbers of subjects, and percentages who died during 5 years of followup. Adherers take 80% or more of prescription.

	<i>Clofibrate</i>		<i>Placebo</i>	
	<i>Number</i>	<i>Deaths</i>	<i>Number</i>	<i>Deaths</i>
Adherers	708	15%	1,813	15%
Non-adherers	357	25%	882	28%
Total group	1,103	20%	2,789	21%

Note: Data on adherence missing for 38 subjects in the clofibrate group and 94 in the placebo group. Deaths from all causes.

Source: The Coronary Drug Project Research Group, "Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project," *New England Journal of Medicine* vol. 303 (1980) pp. 1038–41.

(\* ) Conclusion: Adherers differ from non-adherers in ways that lead to lower mortality among adherers.



## *DEMAND AN EXPERIENCED SURGEON*

*The more experienced a doctor is, the better. As obvious as that sounds, there are still too many people out there who never ask their surgeons for a history of their work. The importance of knowing is illustrated by this study Peter Starek, a surgeon at the University of North Carolina, reviewed 460 heart valve replacement operations and found that only 4 percent of the patients of the three most senior surgeons died. But one junior surgeon lost almost a third of his patients. Since that surgeon was technically the best in the group, says Starek, something was obviously lacking—perhaps the kind of good judgement that grows out of experience...*

From the column of Dr. Dean Edell in the San Francisco Chronicle, 8/1/90.

(\* ) What's missing?

(\* ) Confounding variable – Pre-operation health of the patients upon which each surgeon performed the surgery. Younger, technically better surgeon took the more difficult cases.

## Cervical cancer and oral contraceptives.

A study done of patients at Kaiser Permanente in Walnut Creek, CA, between 1969 and 1971 found that users of oral contraceptives had a higher rate of cervical cancer than non-users.

(\*) Observational study.

(\*) Possible confounding variables: age, education, marital status, religion, smoking.

(\*) Researchers *controlled* for these variables. They divided the data into groups that were homogeneous with respect to these variables and compared rates of cervical cancer between the women taking oral contraceptives and those that weren't in each group separately.

(\*) Same result. Their conclusion — the use of oral contraceptives is a risk factor for cervical cancer.

(\*) They missed one, very important confounding variable....

(\*) Sexual activity.

## Data Collection: Surveys

- (\*) Data is collected from a *sample* of a population.
- (\*) Goal: to learn about the population from the sample.
- (\*) No treatment...
- (\*) ... so no treatment or control groups.
- (\*) Key question: *How well does the sample represent the whole?*