

# Observational studies

---

## Controlled experiment:

The investigator chooses who receives the treatment.

## Observational study:

The subjects themselves choose whether they receive the treatment.

## Key issues:

- Correlation (association) is not causation.
- Confounding
- Simpson's paradox

## The clofibrate trial

---

**Coronary drug project:** A randomized, controlled, double-blind experiment to evaluate 5 drugs (including clofibrate) for the prevention of heart disease.

**Subjects:** Middle-aged men with heart trouble.

- 5,552 assigned at random to the drug groups
- 2,789 assigned at random to the control group (placebo=lactose)
- Patients followed 5 years.

# Clofibrate

---

- Reduces the level of cholesterol in the blood
- Didn't save any lives in the trial:
  - 20% dead in clofibrate group
  - 21% dead in control group
- Possible cause: **adherence**  
(Many subjects didn't take their medicine.)
- In clofibrate group: [Adherer: took >80% of medicine.]
  - Adherers: 15% mortality rate
  - Non-adherers: 25% mortality rate

## Adherence in the clofibrate trial

---

	Clofibrate		Placebo	
	Number	Deaths	Number	Deaths
Adherers	708	15%	1,813	15%
Non-adherers	357	25%	882	28%
Overall	1,103	20%	2,789	21%

Adherers and non-adherers differ in ways other than whether they took their medicine!

**Conclusion:** Clofibrate has no effect.

## Causes of association

---

Suppose  $A$  is associated with  $B$ .

This may be because:

- $A$  causes  $B$
- $B$  causes  $A$
- $X$  is associated with both  $A$  and  $B$ .

## Confounding

---

In the association of  $A$  and  $B$ ,  $X$  is a **confounder** if it is associated with **both**  $A$  and  $B$ .

$X$  need not be a cause of either  $A$  or  $B$ .

For example, in the consideration of **smoking** and **lung cancer**, a **gene** which

- causes **smoking** but is **not** associated with **lung cancer** is **not** a confounder.
- causes **lung cancer** but is **not** associated with **smoking** is **not** a confounder.
- causes **lung cancer** and **is** associated with **smoking** **is** a confounder.

# Disease P——

---

- First observed in Europe in 18th century; important cause of ill-health, disability and premature death among the very poor. Spread across Europe by the beginning of the 19th century.
- It seemed to hit some villages more than others; many households were spared, but some had disease cases year after year.
- Sanitary conditions in diseased households were primitive. Flies were everywhere.
- One blood-sucking fly (*Simulium*) had the same geographical range as the disease, at least in Europe, and was most active in the spring, when most of the cases developed.

Many epidemiologists concluded the disease was infectious and—like malaria, yellow fever, or typhus—was transmitted from one person to another by insects.

Was this conclusion justified?

## Discussion

---

# Cervical cancer and circumcision

---

- Cervical cancer was for many years one of the most common cancers among women.
- Cervical cancer was quite rare among Jews and Moslems.
- In the 1950's, several investigators concluded that circumcision of the males was the protective factor.

Was this conclusion justified?

## Discussion

---

# Ultrasound and low birthweight

---

- Several experiments on lab animals have shown that ultrasound examinations can cause low birthweight. Is this also true for humans?
- Investigators (at Johns Hopkins) ran an observational study to find out.
- The investigators found a number of confounding variables and adjusted for them. Even so, there was an association: Babies exposed to ultrasound in the womb had lower birthweight, on average, than babies who were not exposed.

Is this evidence that ultrasound causes lower birthweight?

## Discussion

---

# Marriage status and child abuse

---

- A newspaper article reported on an organization who had studied the relationship between abuse of a child and the marriage status of the child's parents.
- It was found that children of single parents were more often abused.
- The investigators concluded (quite strongly) that encouraging marriage would reduce the rate of child abuse.

Is this conclusion justified?

## Discussion

---

# Gene expression

---

- Schadt *et al.*, Nature Genetics, 422:297–302 (2003) reported on genome-wide gene expression in liver tissue and obesity in mice who had been on a high-fat, atherogenic diet for 4 months.
- They identified a set of genes that showed quite strong differences in expression in mice with high subcutaneous fat-pad-mass (FPM) versus those with low FPM.
- One might conclude that the gene expression differences contributed to the obesity differences.

Would such a conclusion be justified?

## Discussion

---

## Controlling for a confounder

---

The problem with observational studies is that subjects differ among themselves in crucial ways besides the treatment.

We deal with this by **controlling for** the confounding variable(s)—we compare smaller, more homogeneous subgroups.

- Anticipate, measure, and control for possible confounders.
- Think about other possible confounders that were not considered.
- Never draw very strong conclusions from a single observational study.

## Sex bias in graduate admissions

---

Observational study on sex bias in graduate admissions at the University of California, Berkeley.

During the study period:

- 44% (of 8,442) men were admitted
- 35% (of 4,321) women were admitted

**Does this indicate a sex bias?**

# Sex bias in graduate admissions

---

For the six largest majors:

Major	Men		Women	
	Number of applicants	Percent admitted	Number of applicants	Percent admitted
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7

Simpson's paradox:

Relationships between variables within subgroups can be **reversed** when the subgroups are combined.

## Summary

---

- In **controlled experiments**, we can ensure (by randomization) that the treatment and control groups are the same, except for the treatment.
- In **observational studies**, the treatment and control groups may differ in crucial ways aside from the treatment.
- We try to identify and control for such **confounding** factors.
- We must always consider the possibility that there are additional, unidentified confounders.
- **Correlation is not causation.**