



**DSC 10**  
**Winter 2021**

## Lecture 2

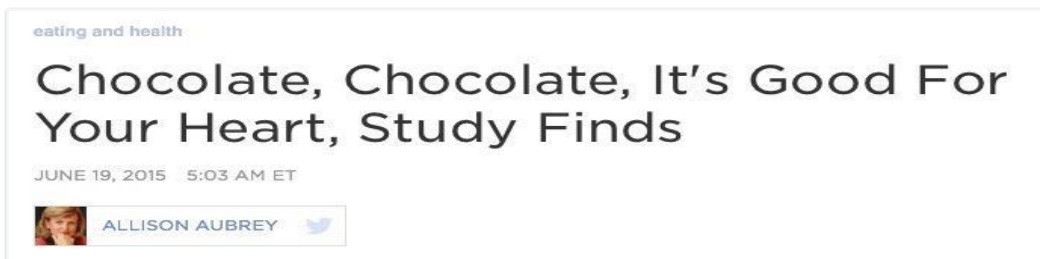
---

Cause & Effect

# **Association and Causality**

# Terminology

- **individuals**, study subjects, participants, units
  - *European adults*
- **treatment**
  - *chocolate consumption*
- **outcome**
  - *heart disease*



[npr.org](http://npr.org) (report on a study in [heart.bmj.com](http://heart.bmj.com))

# The first question: Is there an association?

“Among those in the top tier of chocolate consumption, 12 percent developed or died of cardiovascular disease during the study, compared to 17.4 percent of those who didn’t eat chocolate.”

-Howard LeWine of Harvard Health Blog,  
reported by [npr.org](https://www.npr.org)

- **association**

“any relation”

“link”

# The next question: Is this causal?

---

Does chocolate consumption **lead to** a reduction in heart disease?

- **causality**

This question is often harder to answer.

“[The study] doesn’t prove a cause-and-effect relationship between chocolate and reduced risk of heart disease and stroke.”

– JoAnn Manson, chief of Preventive Medicine at Brigham and Women’s Hospital, Boston

---

**London, 1854**

# Miasmas, miasmaticism, miasmaticists

---

- **Bad smells** given off by waste and rotting matter
  - **Believed to be the main source of disease**
  - Suggested remedies:
    - “fly to cleanse air”
    - “a pocket full o’ posies”
    - “fire off barrels of gunpowder”
  - Staunch believers:
    - Florence Nightingale
    - Edwin Chadwick, Commissioner of the General Board of Health
-

# John Snow, 1813-1858

---





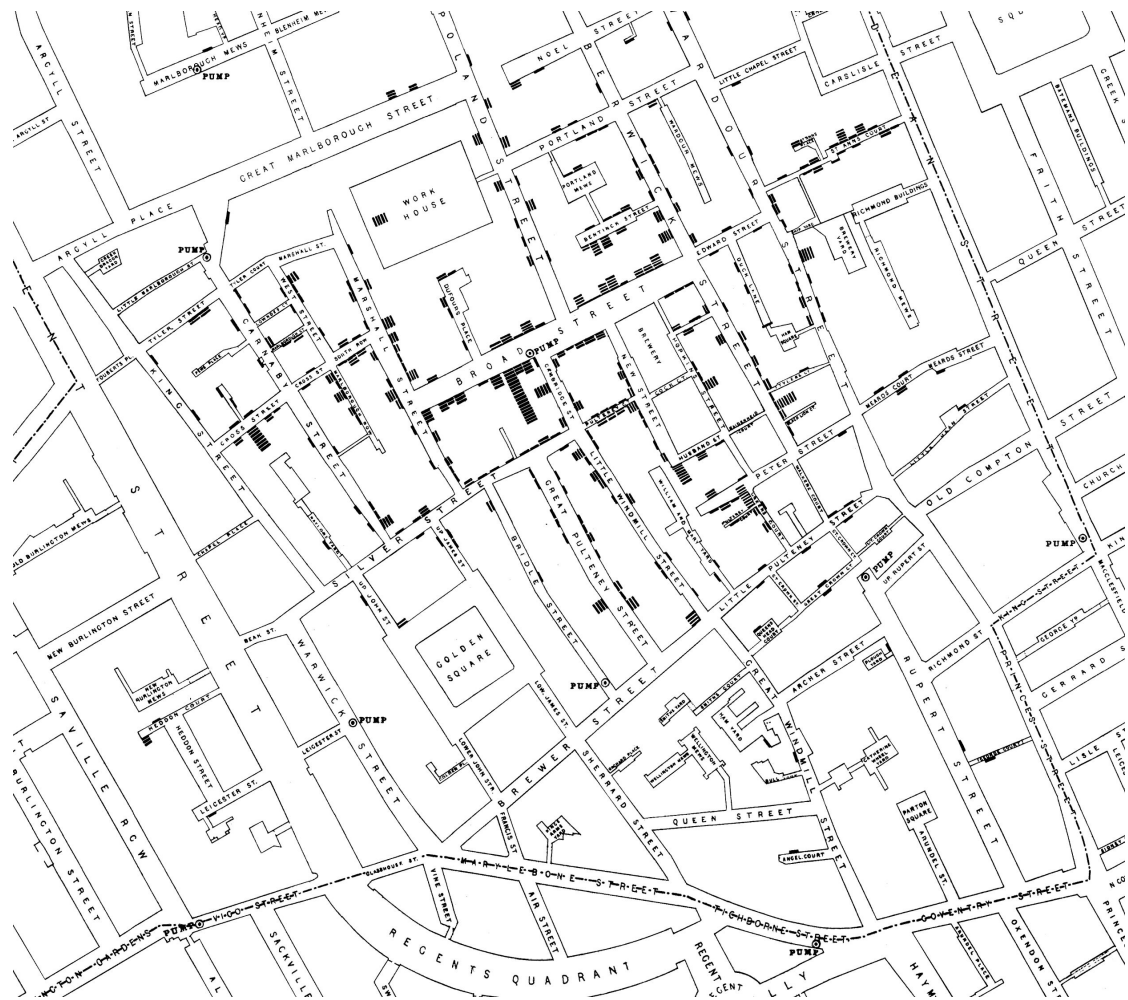
# John Snow, 1813-1858

---



Not this Jon Snow



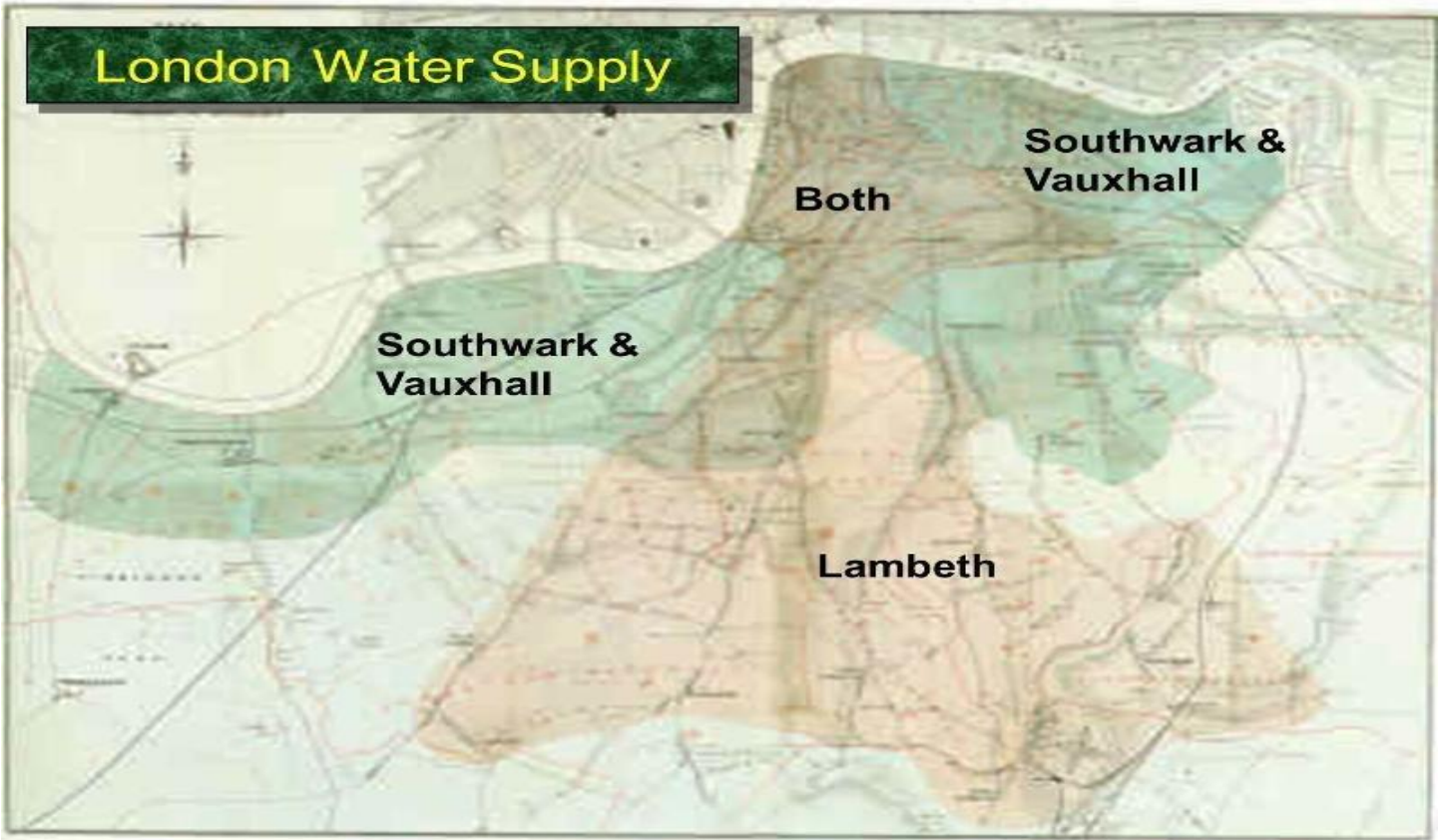




# Establishing Causation



## London Water Supply



# Comparison

---

- **treatment group**
- **control group**
  - does not receive the treatment

Which houses were part of the treatment group?

- A. All houses in the region of overlap
- B. Houses served by S&V (dirty water) in the region of overlap
- C. Houses served by Lambeth (clean water) in the region of overlap

# Snow's “Grand Experiment”

---

“... there is no difference whatever in the houses or the people receiving the supply of the two Water Companies, or in any of the physical conditions with which they are surrounded ...”

- The two groups were *similar except for the treatment*.

# Snow's table

| Supply Area           | Number of houses | Cholera deaths | Deaths per 10,000 houses |
|-----------------------|------------------|----------------|--------------------------|
| S&V (dirty water)     | 40,046           | 1,263          | 315                      |
| Lambeth (clean water) | 26,107           | 98             | 37                       |
| Rest of London        | 256,423          | 1,422          | 59                       |

Does dirty water cause cholera?

- A. Yes, I think so
- B. No, I don't think so
- C. Maybe, I can't tell



# Key to establishing causality

---

If the treatment and control groups are *similar apart from the treatment*, then differences between the outcomes in the two groups can be ascribed to the treatment.

---

# Confounding

# Trouble

---

If the treatment and control groups have **systematic differences other than the treatment**, then it might be difficult to identify causality.

Such differences are often present in **observational studies**.

When they lead researchers astray, they are called **confounding factors**.



# Randomize!

---

- If you assign individuals to treatment and control **at random**, then the two groups are likely to be similar apart from the treatment.
  - You can account – mathematically – for variability in the assignment.
  - **Randomized Controlled Experiment**
-

# Randomized Controlled Experiments

---

- Assign individuals to treatment and control **at random**

Which of these questions cannot be answered by running a randomized controlled experiment?

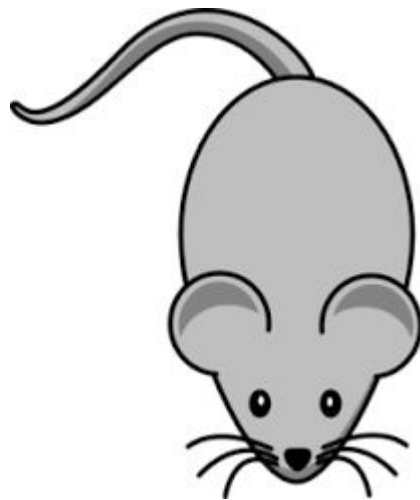
- A. Does daily meditation reduce anxiety?
  - B. Does playing video games increase aggressive behavior?
  - C. Does smoking cigarettes cause weight loss?
  - D. Does early exposure to classical increase a child's IQ?
-

# Careful ...

---

Regardless of what the dictionary says,  
in probability theory

**Random  $\neq$  Haphazard**



# **Summary: Cause & Effect**

# Comparison

---

- Group by some *treatment* and measure some *outcome*
  - Simplest setting: a *treatment group* and a *control group*
  - If the *outcome* differs between these two groups, that's evidence of an *association* (or *relation*)
    - E.g., the top-tier chocolate eaters died of heart disease at a lower rate (12%) than chocolate abstainers (17%)
  - If the two groups are similar in all ways but the *treatment*, a difference in the *outcome* is also evidence of *causality*
-



# Confounding

---

- If the treatment and control groups have systematic differences other than the treatment itself, then it might be difficult to identify a causal link
  - When these systematic differences lead researchers astray, they are called *confounding factors*
  - Such differences are often present in observational studies
    - *Observational study*: the researcher does not choose which subjects receive the treatment
    - *Controlled experiment*: the researcher designs a procedure for selecting the treatment and control groups
-

# Randomize!

---

- When subjects are split up *randomly*, it's unlikely that there will be systematic differences between the groups
  - And it's possible to account for the chance of a difference
  - Therefore, *randomized controlled experiments* are the most reliable way to establish causal relations
-