Why Study Causality?

- Machinery developed so far (OLS) applicable to any population model, whether causal or not
  - Doesn't care about changes to a specific *x*
  - Operates on datasets where the *x* values are fixed
- Often desirable to measure causal relationships
- Goal of estimating a model: make a decision or effect a change
  - Importance of counterfactual
    - Can't be observed, can only measure correlations

Correlation Is Not Causation: Example 1

Is drinking red wine good for your heart?

- Many studies associate drinking wine with lower risk of heart disease.
  - Problem: observational
- Many reasons might explain why wine drinkers have lower risk.
  - Rest of diet
  - Leisure/family time

Drinking wine is an endogenous variable; it's correlated with other factors that also affect risk of heart disease.

Correlation Is Not Causation: Example 2

Is pollution bad for your health?

- Observational studies associate:
  - Higher scores on health measures with people who live in cleaner areas.
  - Lower scores with people who live in polluted areas.
- There are differences between people who live in clean vs. polluted areas.
  - Wealth
  - Diet, stress, and exercise

Pollution exposure is an endogenous variable; it's correlated with many other factors that affect health.

Correlation Is Not Causation: Example 3

Does buying umbrellas cause rain?

- Strong correlation between the two exists.

Buying umbrellas is an endogenous variable.

- It's correlated with many atmospheric factors that could affect rainfall.

Randomized Control Trials (RCTs)

- Developed in medical field, in large part to deal with endogeneity issues
- Makes treatment variable truly random
  - Uncorrelated with any other random variable
  - Exogenous
- Guarantees that any unmeasured variable is the same, in expectation, for treatment and control groups
  - Measured effects truly causal
- First appeared in medical journal in 1948
- In 1970, became part of official FDA regulations

"Intuitively, the causal effect of one treatment, E, over another, C, for a particular unit and an interval of time from t_1 to t_2 is the difference between what would have happened at time t_2 if the unit had been exposed to E initiated at t_1 and what would have happened at t_2 if the unit had been exposed to C initiated at t_1 : 'If an hour ago I had taken two aspirins instead of just a glass of water, my headache would now be gone,' or, 'because an hour ago I took two aspirins instead of just a glass of water, my headache is now gone.'"

—Donald Rubin

Causal Frameworks for Social Science

- Migrating from medical literature into social sciences by early 1980s
- Donald Rubin (1970s): Rubin causal model
  - Based on idea that treatment vs. control corresponds to two counterfactual states of the world
  - Requires that everything else stays the same other than the decision made
- Possible to measure causal effects for decisions we can treat as random