

“We’re All Social Scientists Now”

Introduction to Quantitative Social Science

Do Police Body Cameras Affect the Use of Force?



Do Police Body Cameras **Affect** the Use of Force?

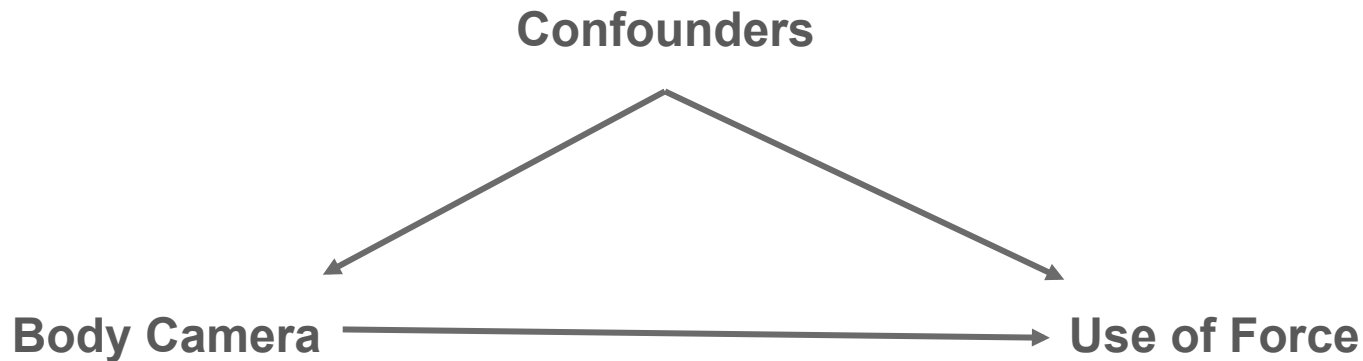


Do Police Body Cameras Affect the Use of Force?



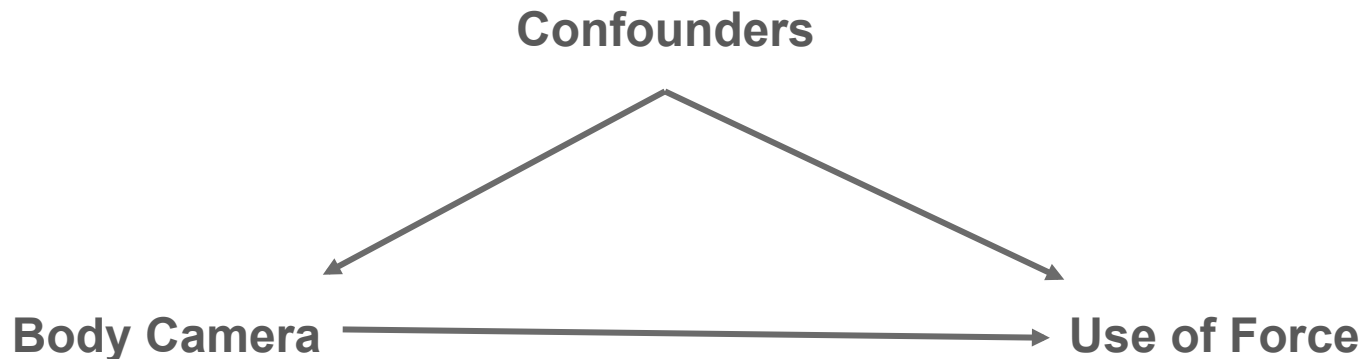
Naive estimator: $E[\# \text{ uof} \mid \text{body cam}] - E[\# \text{ uof} \mid \sim \text{body cam}]$

Do Police Body Cameras Affect the Use of Force?



- Examples?

Do Police Body Cameras Affect the Use of Force?

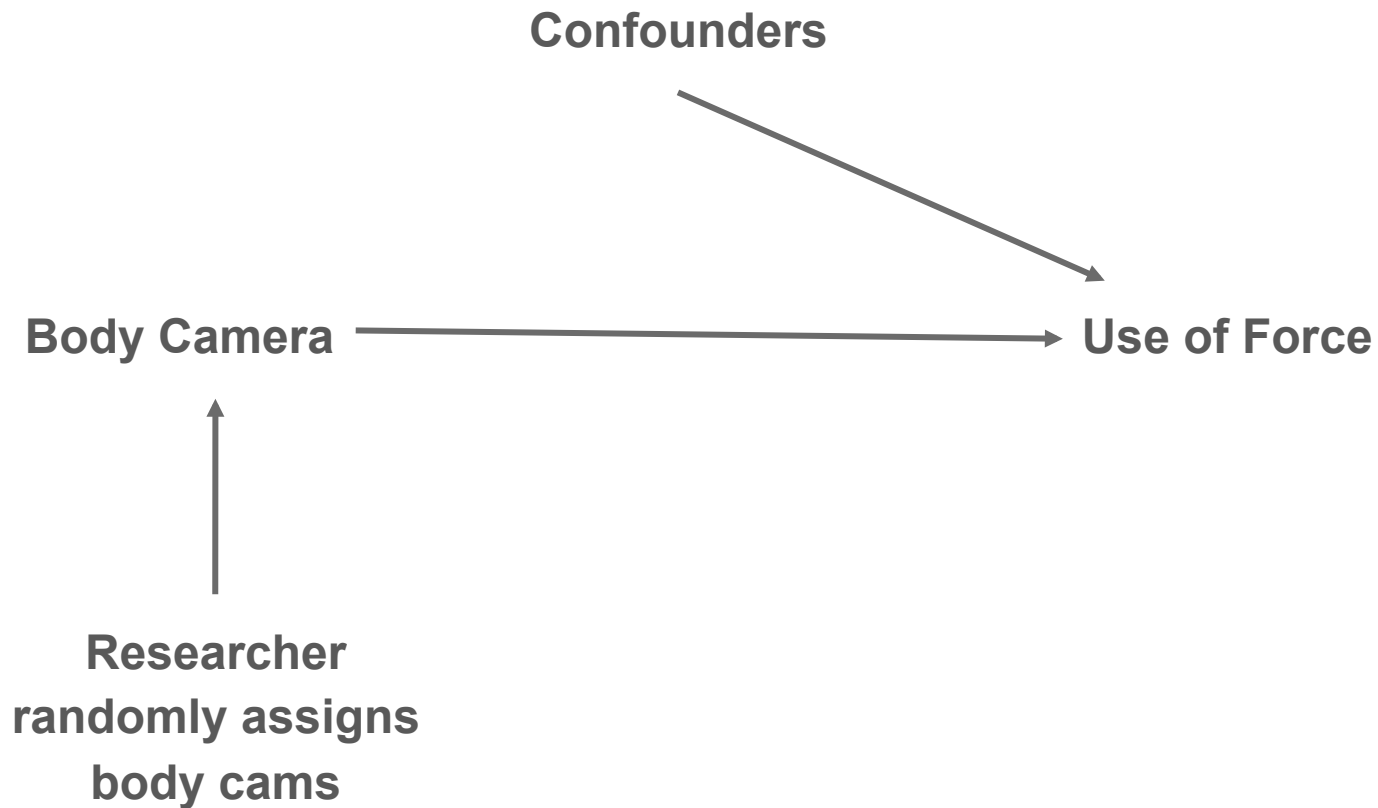


- Examples?
 - Consent decree, union, dept policy, officer assignment (e.g. undercover and patrol danger), officer quality, gvt level (fed, state, local), race, age, gender, economic status

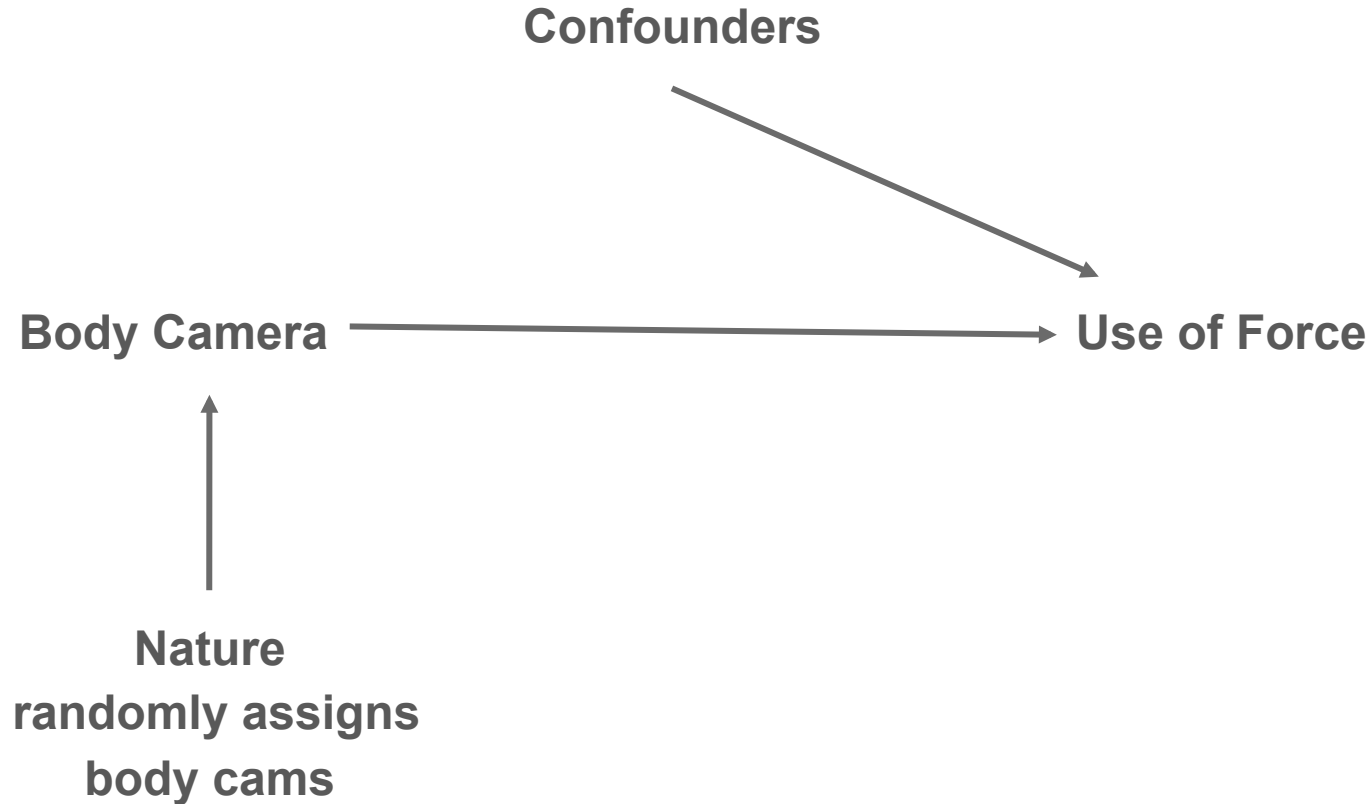
Causal Research Design

1. Experiments
2. Quasi-experiments
3. Observational studies

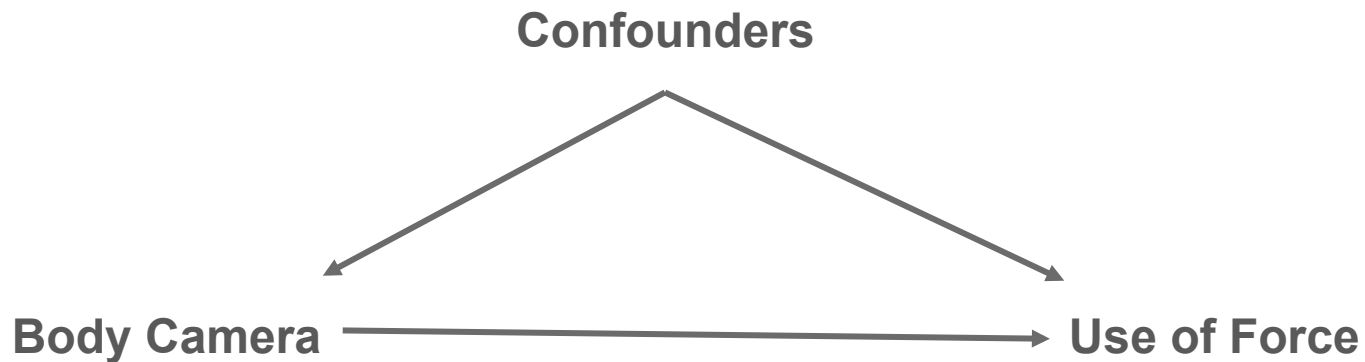
Experiment



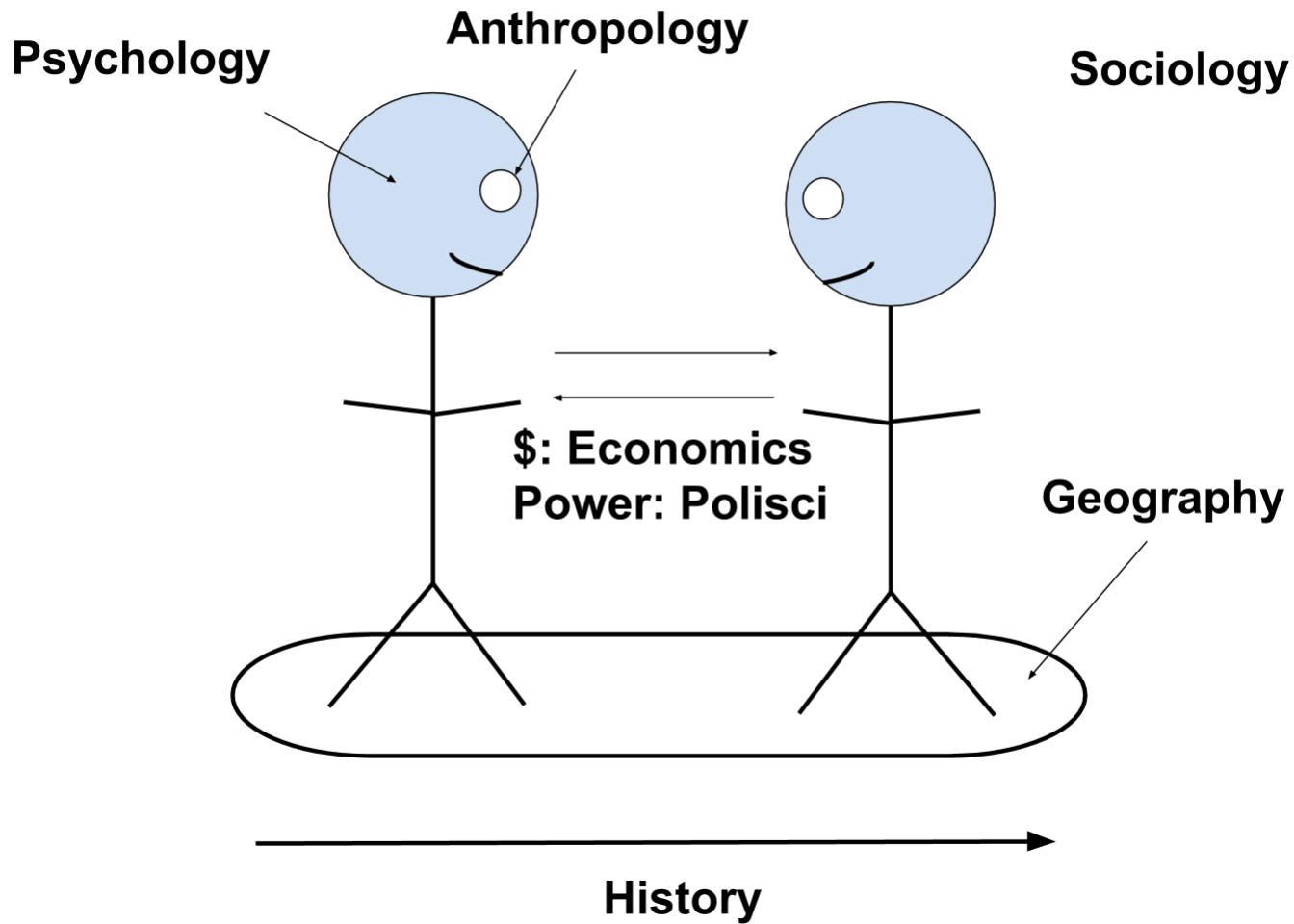
Quasi-Experiment



Observational Study



Include confounders in regression



Why Social Science?

- We're interested in people
- We want to change behaviors
- Privacy, bias, inclusion

Why Social Science?

- We're interested in people
- We want to change behaviors
- Privacy, bias, inclusion

Social scientists have been thinking about these things for a long time

Why Social Science?

- We're interested in people
- We want to change behaviors
- Privacy, bias, inclusion

Social scientists have been thinking about these things for a long time

“Statistics”: from “state.” It was meant to improve government.

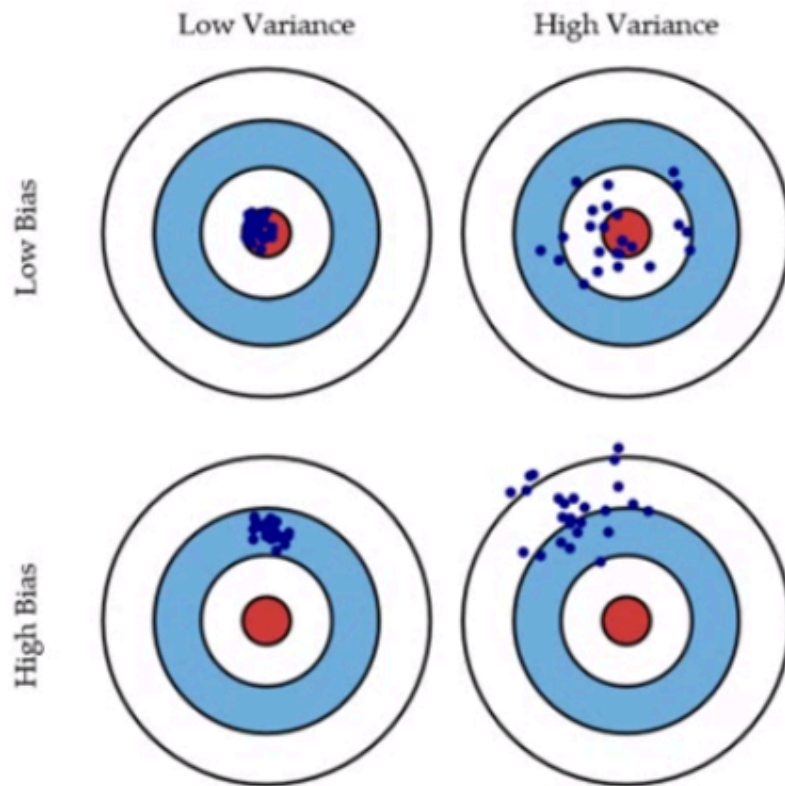
Machine Learning vs Social Science

	Machine Learning	Social Science
Primary Goal	Prediction	Explanation
Model Selection	Out-of-sample accuracy	Theory & Parameter Estimates

Body Camera Example

- Machine learning example:
 $E[\# \text{ uof} \mid \text{bodycam, confounders \& everything else}]$
- Social science example:
 $E[\# \text{ uof} \mid \text{bodycam, confounders}] - E[\# \text{ uof} \mid \sim \text{bodycam, confounders}]$

Machine Learning vs Social Science



Machine Learning vs Social Science

Target A
Poor Validity,
Good Reliability



Target B
Poor Validity
Poor Reliability



Target C
Good Validity,
Good Reliability



Machine Learning vs Social Science

Target A
Poor Validity,
Good Reliability



Target B
Poor Validity
Poor Reliability



Target C
Good Validity,
Good Reliability



Reliability sets the upper bound on validity

Important Social Science Issues

- Too few rows
 - No counterfactuals
 - External validity: E.g. studying self-control fatigue in the lab
 - Selection effect: E.g. vulnerable populations are less likely to report police misconduct
 - Ecological inference
- Too many rows
 - p hacking, in-sample fitting and testing: E.g., flexible post-hoc outlier exclusion
- Too few columns
 - Unobserved variables: E.g. 2013 NFP project: motivation of mothers in/out of the program
- Too many columns
 - Post-treatment controls
 - Noise
- The wrong values
 - Missing values
 - Noisy values (e.g. conceptualization - operationalization - measurement)
 - Systematically biased values
- Model specification
 - Relationship between rows, columns, values, e.g. linear v. non-linear, SUTVA
- Ethics
 - E.g. send mailers placing non-partisan judicial candidates with Montana state seal

Some Lessons Learned for DSSG Projects

- Behaviors are better predictors than demographics
- Some important predictors
 - Race
 - Age
 - Gender
 - Econ status
 - Geography
- Some useful datasets
 - American Community Survey
 - American Time Use Survey
 - General Social Survey
 - Behavioral Risk Factor Surveillance System
- Model all people involved (politicians, inspectors, judges, not just defendants)
- Potential bias at every step

Social Science Examples

- Miller's Law: 7 +/- 2
- Culture of Honor
- Invisible Gorilla: Selective Attention
- Implicit Association Test
- Social Pressure and Voter Turnout
- The Michigan Model (partisan ID)
- Republicans Should Pray for Rain
- The Political Legacy of American Slavery
- Effective Messages in Vaccine Promotion
- Crime and Punishment
- Geography and Trade
- Political Economy of Terrorism
- Minimum Wages and Employment