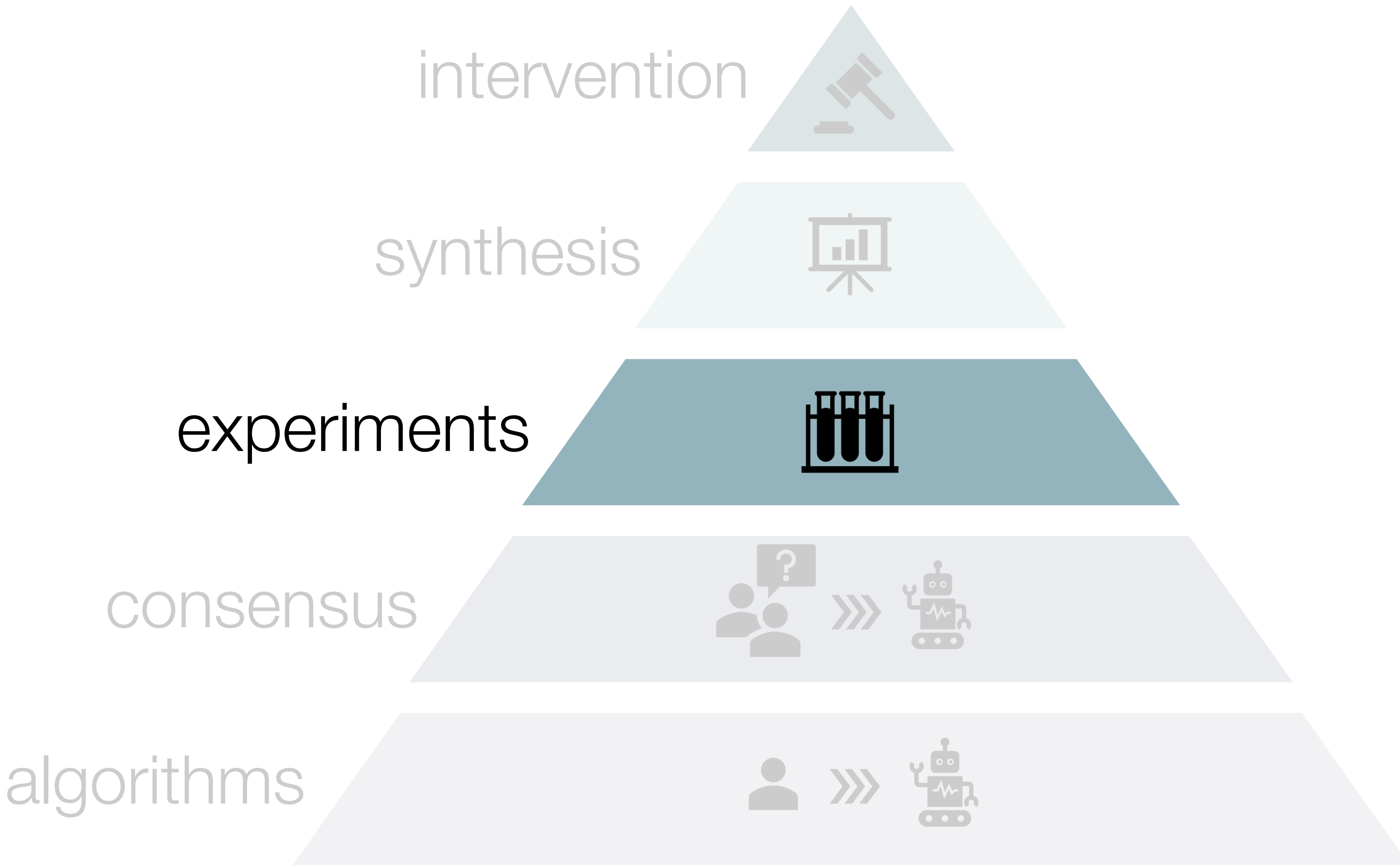
# Evidence-based Decision Making
## Counterfactuals: Alternatives to Experiments

Loreen Tisdall, FS 2024

Version: April 15, 2024


intervention
synthesis
experiments
consensus
algorithms

# Goals for today

- Understand the nature of causal inference as the comparison of treatment to some counterfactual

- Familiarize yourself with different methods of causal inference - randomization, regression, regression discontinuity, instrumental variables - and associated limitations

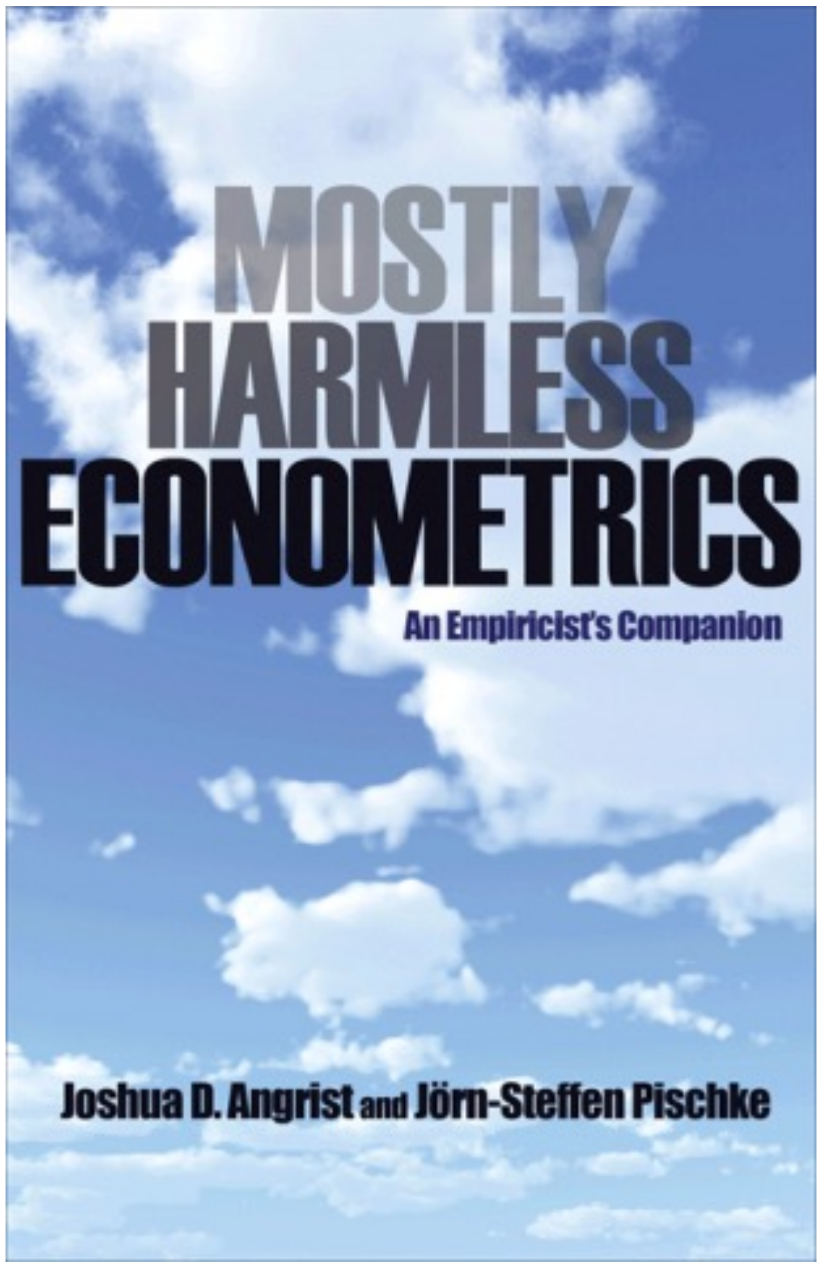
- Consider the untapped potential of natural experiments

# A colorful bouquet of creating counterfactuals

“The stronger the demonstrated consistency of an association under conditions that rule out alternative hypotheses and the stronger the evidence regarding a mechanism that can explain the observed association, the more likely we are to accept the causal hypothesis. Usually the evidence required to confirm a causal hypothesis is cumulated across multiple studies, many of which are, of necessity, observational. Although a wide variety of research designs and analytic techniques are available to assist in gathering evidence to support a causal inference, they are helpful only to the extent that their use is guided and constrained by appropriate subject-matter considerations. **No method or set of methods defines causality**.”

Marini, M. M., & Singer, B. (1988). Causality in the social sciences. *Sociological methodology, 18*, 347-409.

# "Furious Five" statistical methods for causal inference




- Randomisation
- Difference in differences
- Regression
- Regression discontinuity
- Instrumental variables

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. http://doi.org/10.1257/jep.24.2.3

# “Furious Five” statistical methods for causal inference

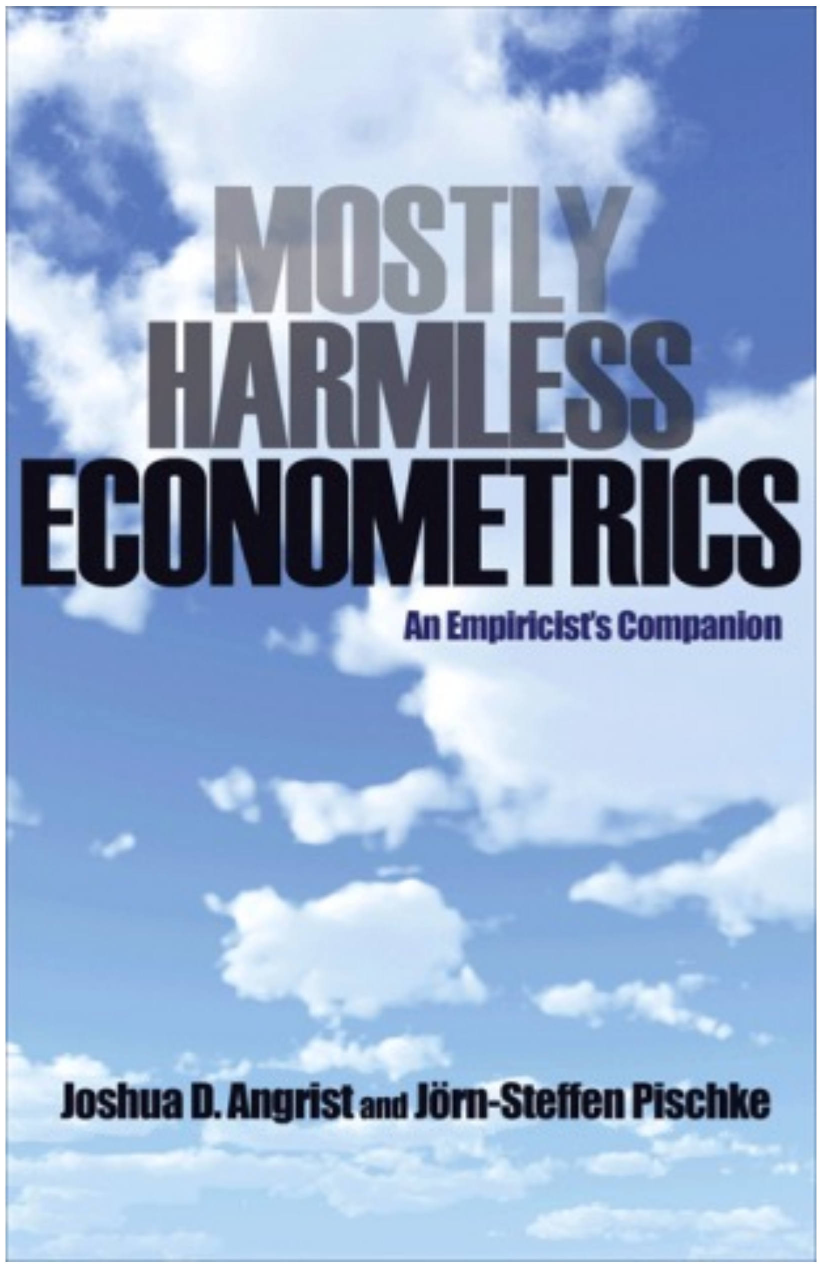
Image created with AI (Bing), February 13, 2024

YOUR TURN!

Take 1 minute and write down what you think is the key idea behind each of the “Furious Five”.

1) Randomisation,
2) Difference in differences,
3) Regression,
4) Regression discontinuity,
5) Instrumental variables

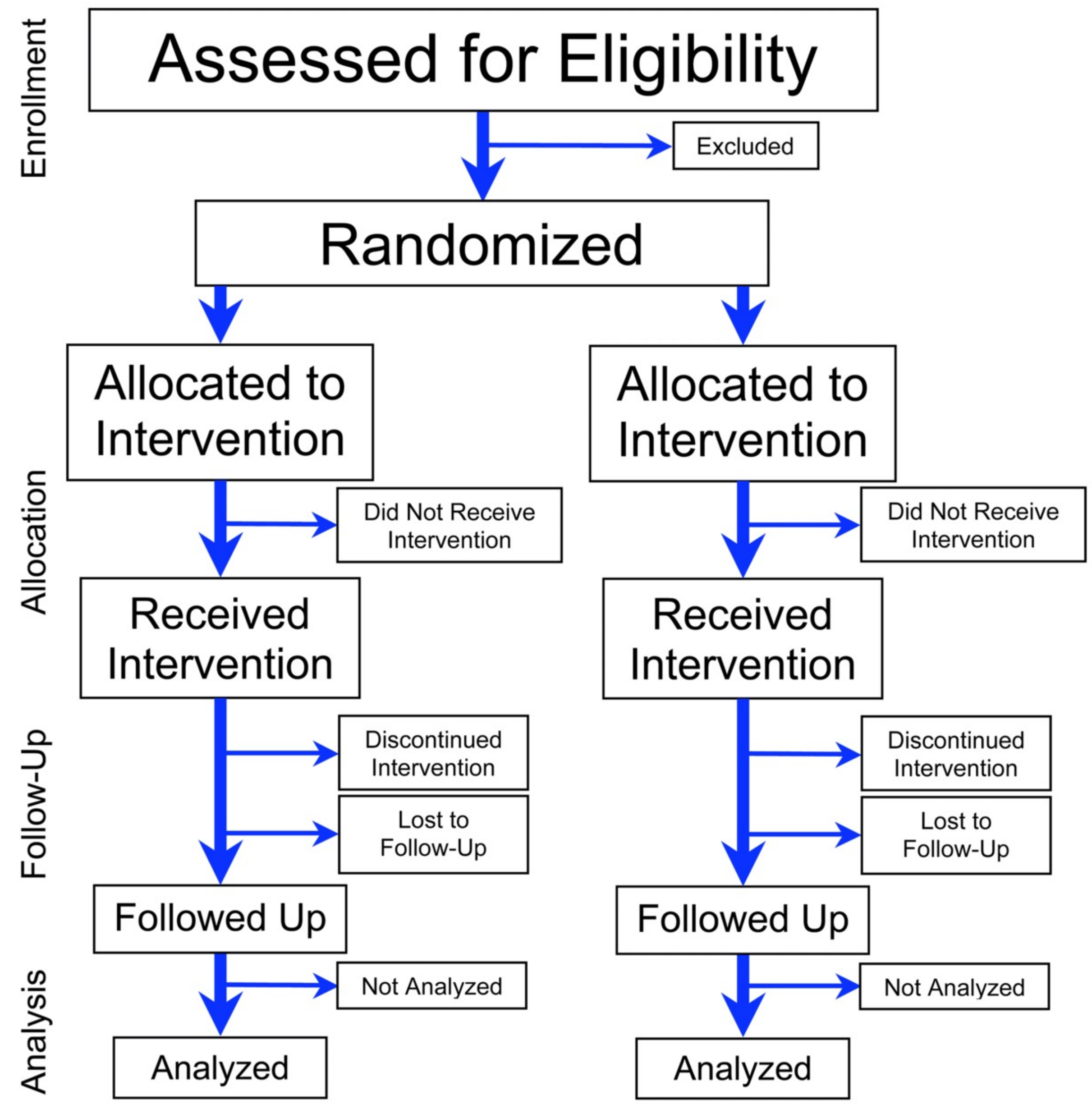
# “Furious Five” statistical methods for causal inference




- **Randomisation**
- Difference in differences
- Regression
- Regression discontinuity
- Instrumental variables

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. http://doi.org/10.1257/jep.24.2.3

# Randomisation

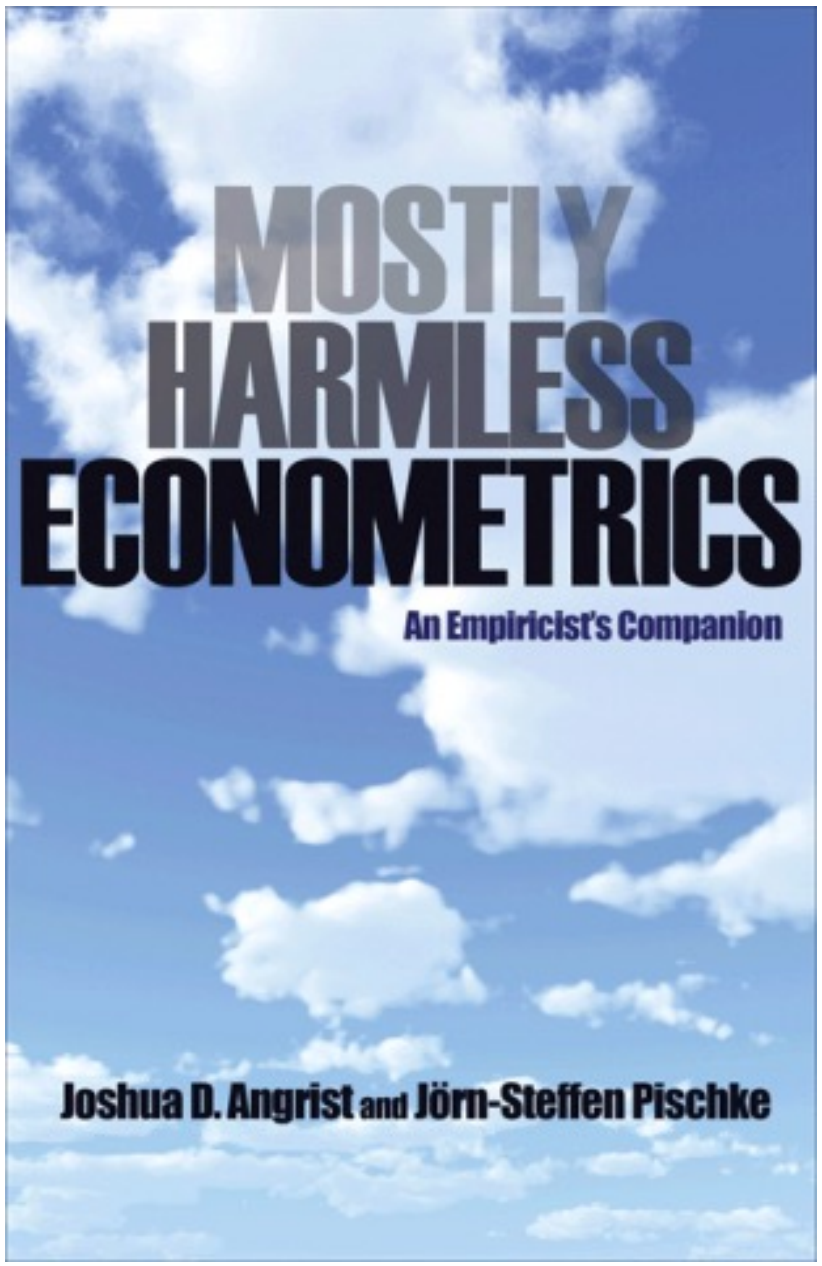


# Full randomisation is seldom available in practice…

The "ideal" data, from the viewpoint of the analyst, would be data from an incompetent advertiser who allocated expenditures randomly across cities. If ad expenditure is truly random, then we do not have to worry about confounding variables because the predictors will automatically be orthogonal to the error term. However, statisticians are seldom lucky enough to have a totally incompetent client.

Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

# “Furious Five” statistical methods for causal inference



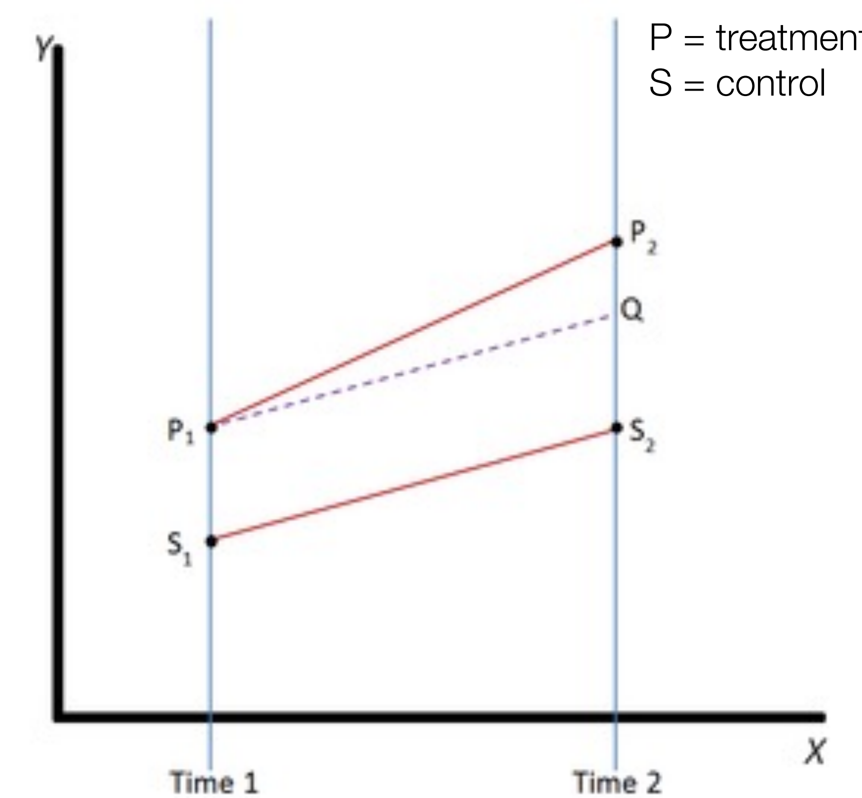

- Randomisation
- **Difference in differences**
- Regression
- Regression discontinuity
- Instrumental variables

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. http://doi.org/10.1257/jep.24.2.3

# Difference in differences

Difference in differences (DID or DD) is a statistical technique used in the social sciences that attempts to mimic an experimental research design using observational study data, by studying the differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment.

It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group. Although it is intended to mitigate the effects of extraneous factors and selection bias, depending on how the treatment group is chosen, this method may still be subject to certain biases (e.g., omitted variable bias).



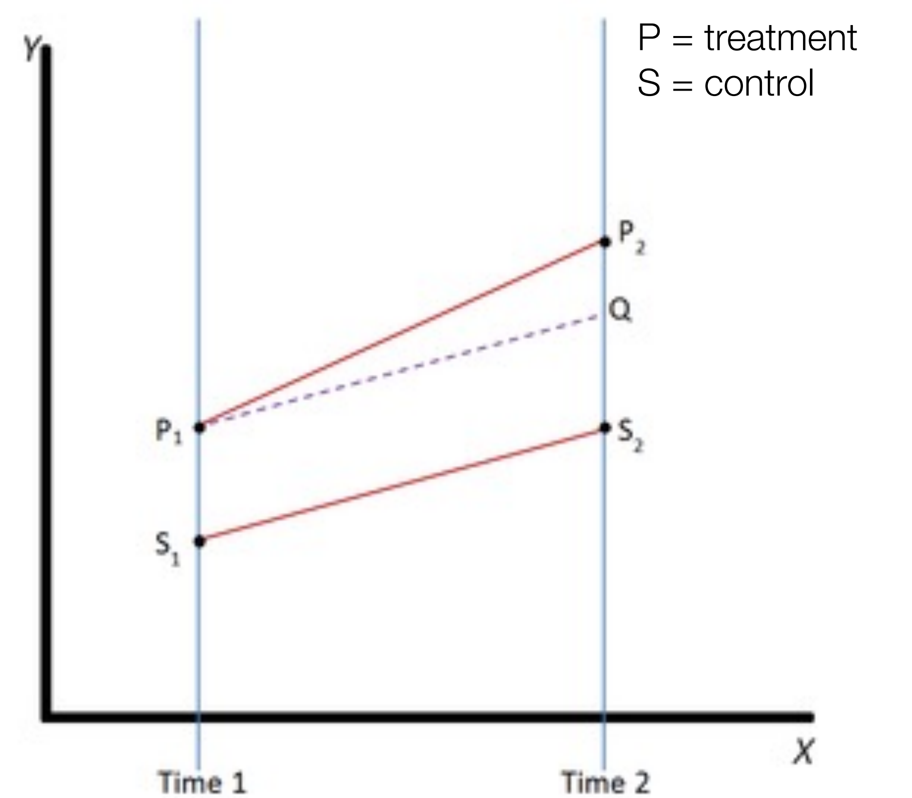

$$Y = [P2 - P1] - [S2 - S1]$$

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics, 119*(1), 249–275.

# Difference in differences

Difference in differences (DID or DD) is a statistical technique used in the social sciences that attempts to mimic an experimental research design using observational study data, by studying the differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment.

It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group. Although it is intended to mitigate the effects of extraneous factors and selection bias, depending on how the treatment group is chosen, this method may still be subject to certain biases (e.g., omitted variable bias).



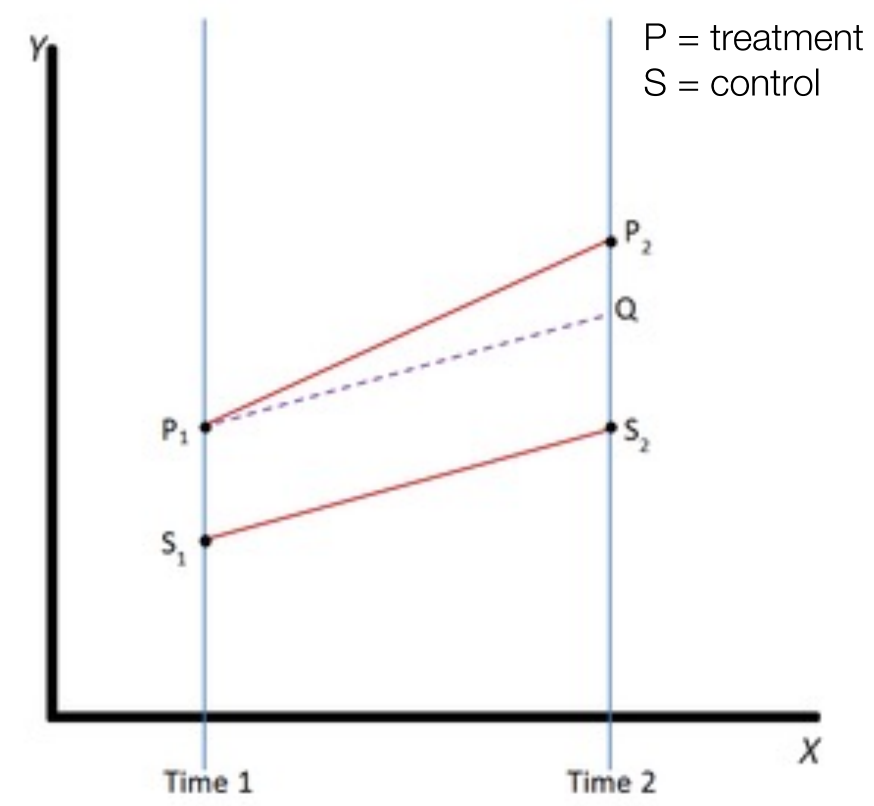

$$Y = B_0 + B_1 Group + B_2 Time + B_3\, Group*Time$$

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics, 119*(1), 249–275.

# Difference in differences

Difference in differences (DID or DD) is a statistical technique used in the social sciences that attempts to mimic an experimental research design using observational study data, by studying the differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment.

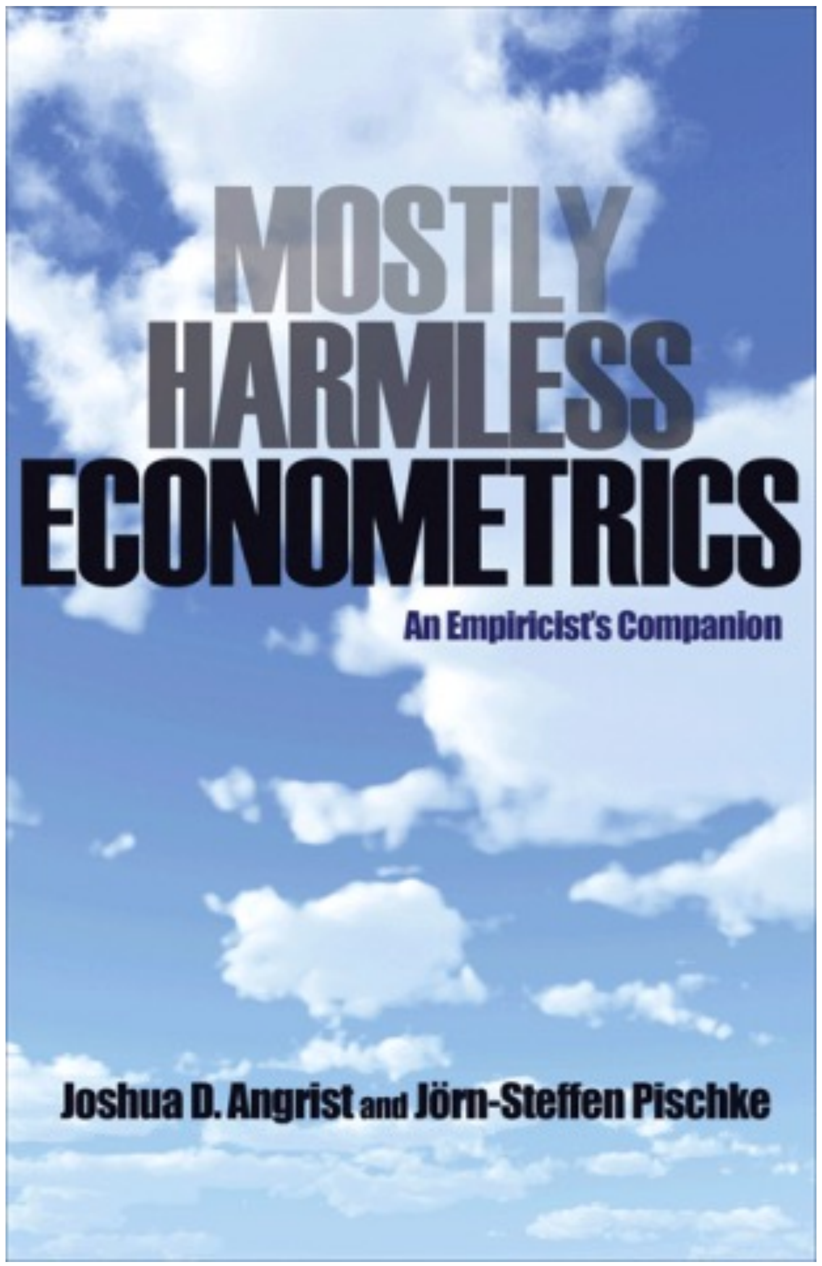
It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group. Although it is intended to mitigate the effects of extraneous factors and selection bias, depending on how the treatment group is chosen, this method may still be subject to certain biases (e.g., omitted variable bias).




**Problem**: Assumption that the change in outcomes from pre- to post-intervention in the control group (S) is a good proxy for the (counterfactual) change in untreated potential outcomes in the treated group (P) may not be warranted; choice of treatment/control groups is crucial (an additional trick may be *matching* on observables)…


Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics, 119*(1), 249–275.

# “Furious Five” statistical methods for causal inference




- Randomisation
- Difference in differences
- **Regression**
- Regression discontinuity
- Instrumental variables

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. http://doi.org/10.1257/jep.24.2.3

# Regression

Regression analysis is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable (criterion) and one or more independent variables (predictors). More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are fixed.

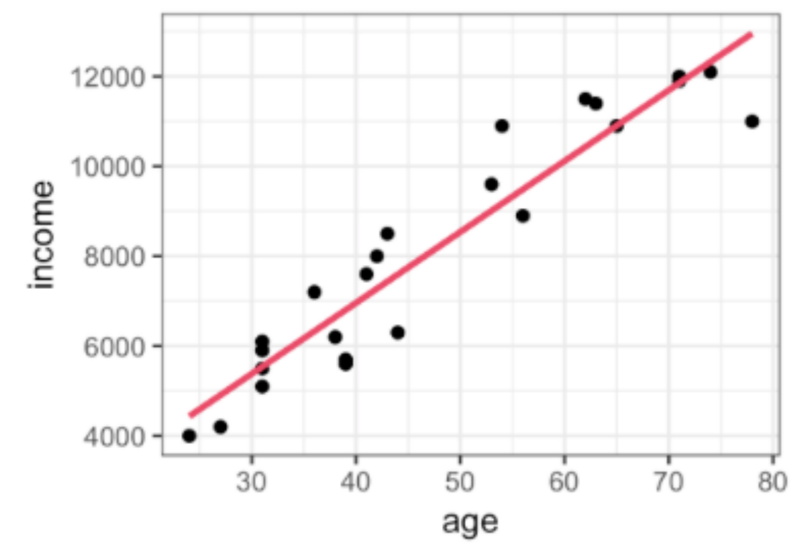
# Regression

## Simple Linear Regression

**Definition**: Simple linear regression is a linear model with one predictor $x$, and where the error term $\epsilon$ is Normally distributed.

$$y = \beta_0 + \beta_1 x + \epsilon$$

| Parameter | Description | In words |
|---|---|---|
| $\beta_0$ | Intercept | When x = 0, what is the predicted value for y? |
| $\beta_1$ | Coefficient for x | For every increase of 1 in x, how does y change? |




Formula

$$income = 885 + 149.3 \times age + \epsilon$$

Coefficients

$$\beta_0 = 885, \beta_{age} = 149.3$$

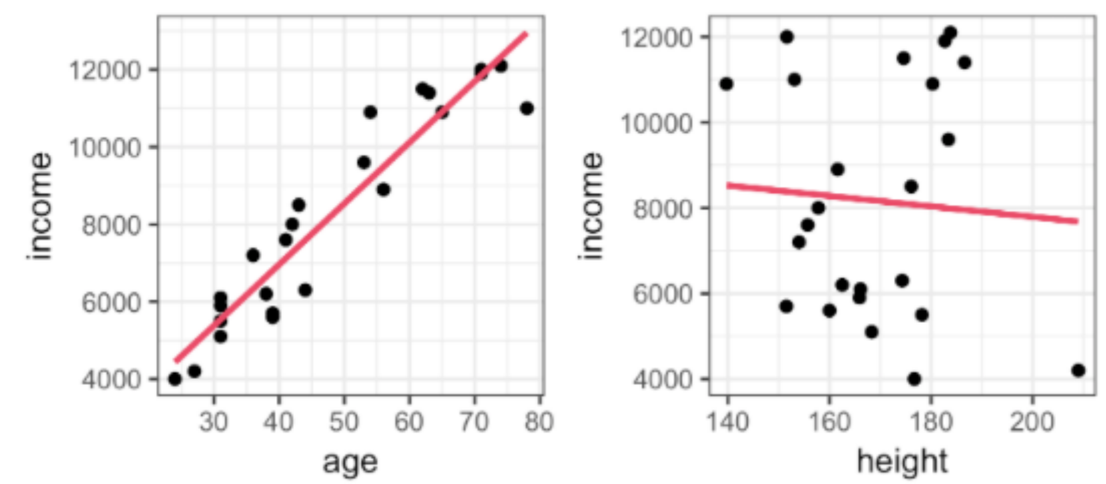
# Regression

## Multiple Linear Regression

**Definition**: Multiple linear regression is a linear model with many predictors $x_1, x_2, \ldots. x_n$, and where the error term $\epsilon$ is Normally distributed.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

| Parameter | Description | In words |
|---|---|---|
| $\beta_0$ | Intercept | When all x values are 0, what is the predicted value for y? |
| $\beta_1, \beta_2, \ldots$ | Coefficient for $x_1, x_2, \ldots$ | For every increase of 1 in coefficient for $x_1, x_2, \ldots$ how does y change? |



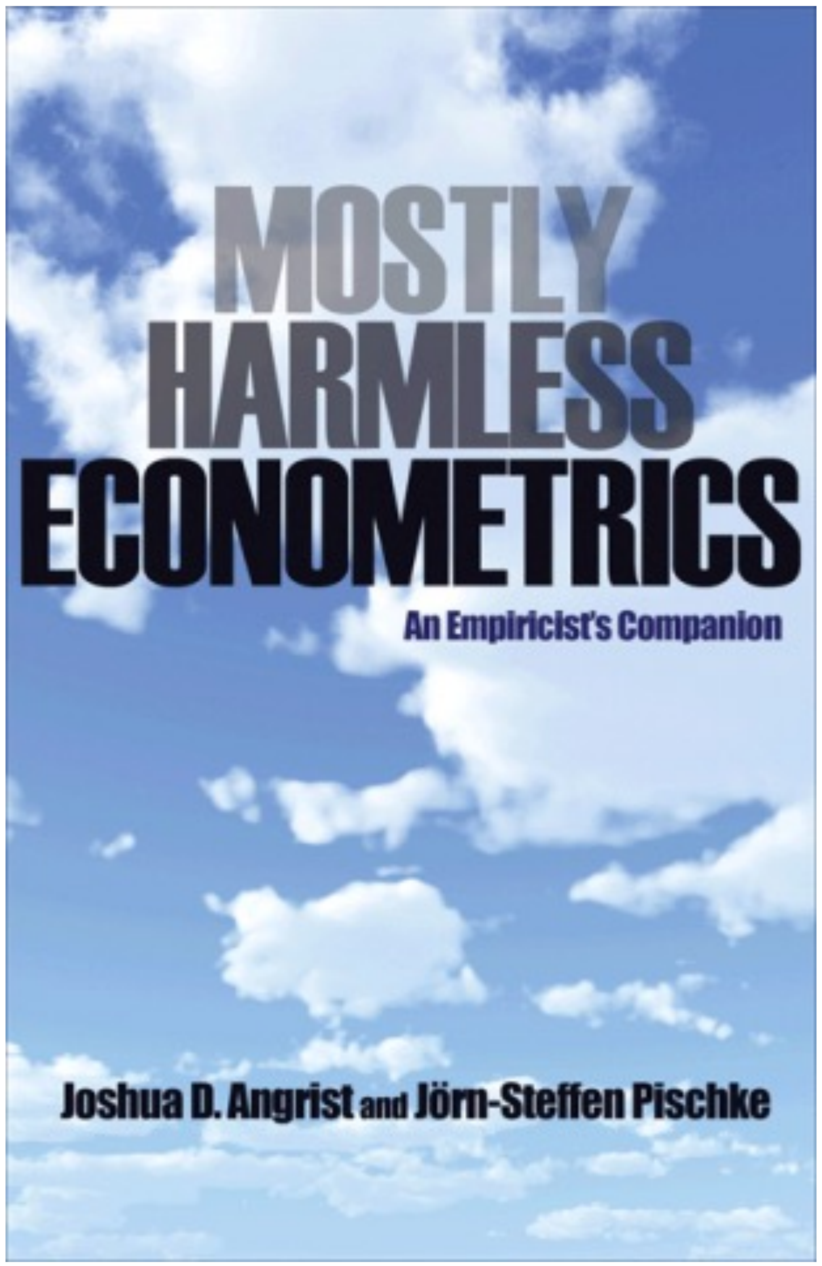

Formula

$$income = 1628 + 147 \times age - 4.1 \times height + \epsilon$$

Coefficients

$$\beta_0 = 1628, \beta_{age} = 147, \beta_{weight} = -4.1$$

# “Furious Five” statistical methods for causal inference

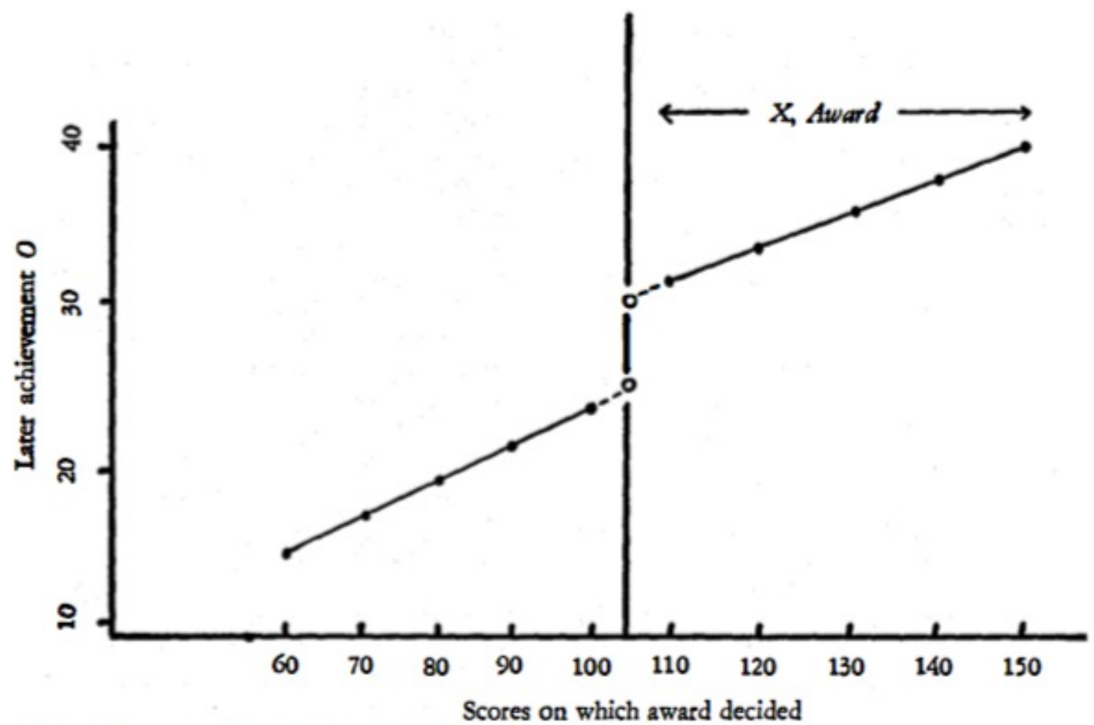



- Randomisation
- Difference in differences
- Regression
- **Regression discontinuity**
- Instrumental variables

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. http://doi.org/10.1257/jep.24.2.3

# Regression discontinuity

A regression discontinuity design (RDD) is a quasi-experimental pretest-posttest design that elicits the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned. By comparing observations lying closely on either side of the threshold, it is possible to estimate the average treatment effect in environments in which randomization is unfeasible. RDD was first applied by Donald Thistlethwaite and Donald Campbell to the evaluation of scholarship programs.



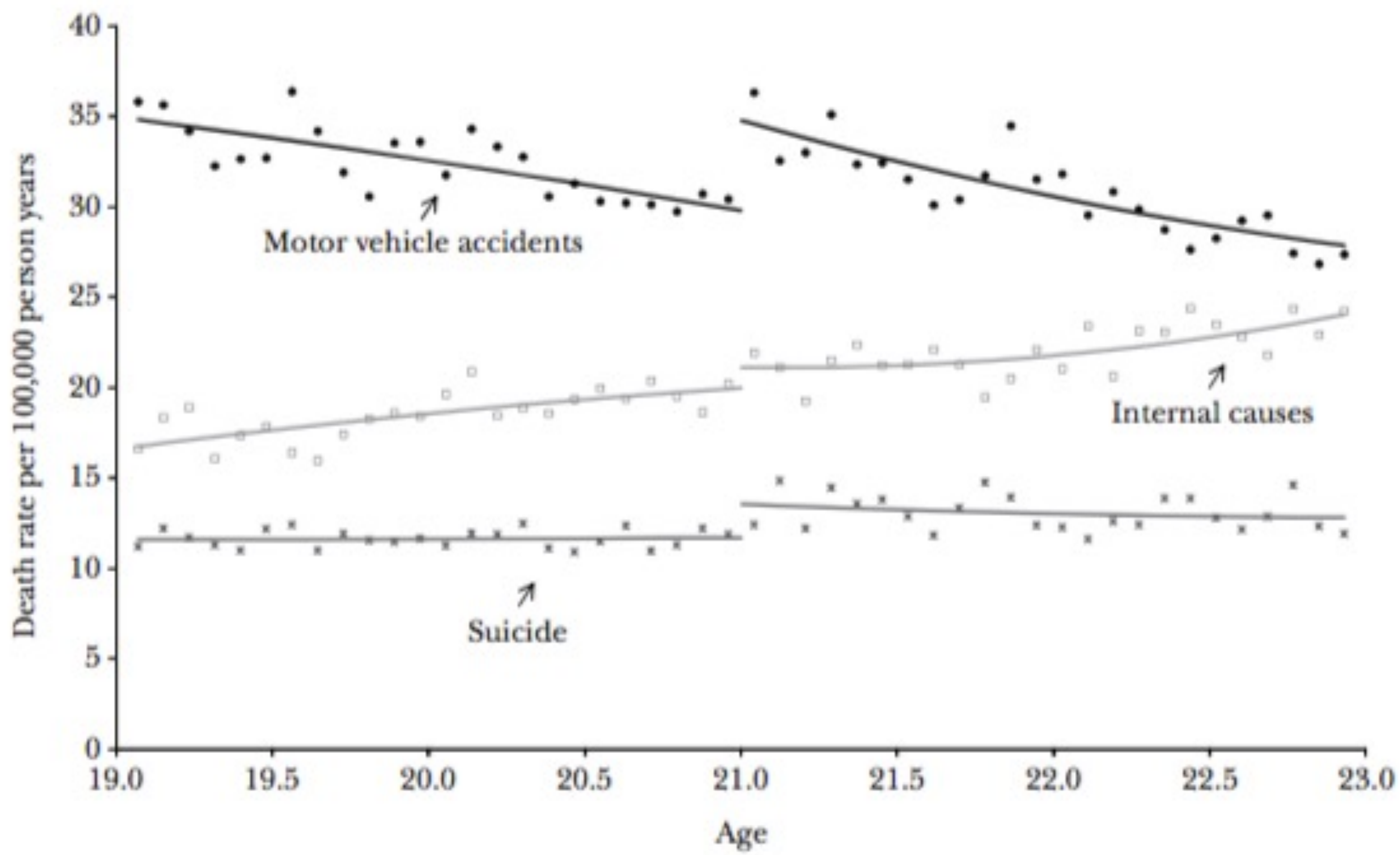

$$Y = B_0 + B_1\text{Score} + B_2\,\text{Award}$$

**Problem**: Assumption that the individuals just below the cutoff are not systematically different from those just above can be wrong (e.g., individuals just above the threshold could try harder); the estimation may not generalise to observations away from the cutoff (e.g., awards could have different results at different levels of ability).

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature, 48*(2), 281–355. http://doi.org/10.1257/jel.48.2.281

# Regression discontinuity

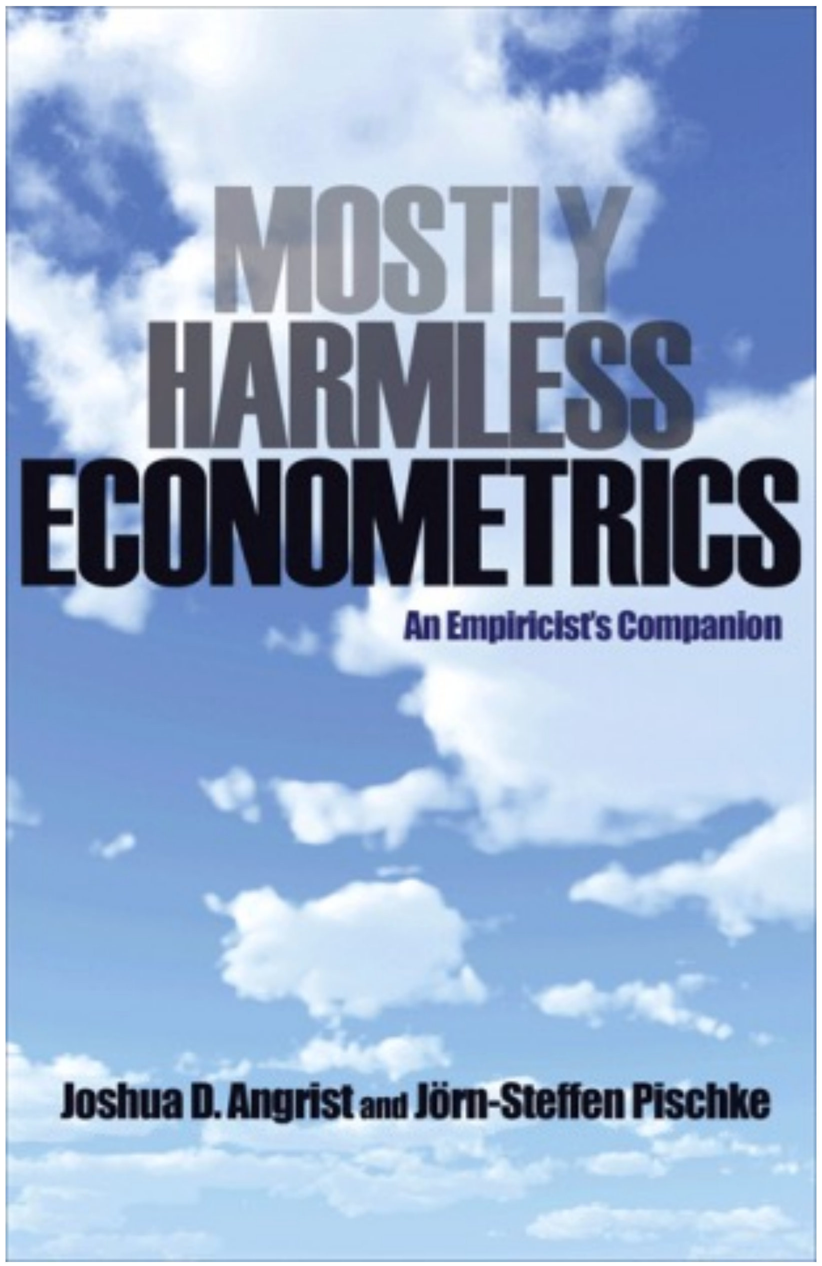
*Figure 2*
**Age Profiles for Death Rates in the United States**




*Notes:* The death rates are estimated by combining the National Vital Statistics records with population estimates from the U.S. Census.

Carpenter, C., & Dobkin, C. (2011). The Minimum Legal Drinking Age and Public Health. *Journal of Economic Perspectives*, *25*(2), 133–156.

# "Furious Five" statistical methods for causal inference




- Randomisation
- Difference in differences
- Regression
- Regression discontinuity
- **Instrumental variables**

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. http://doi.org/10.1257/jep.24.2.3

# Instrumental variables

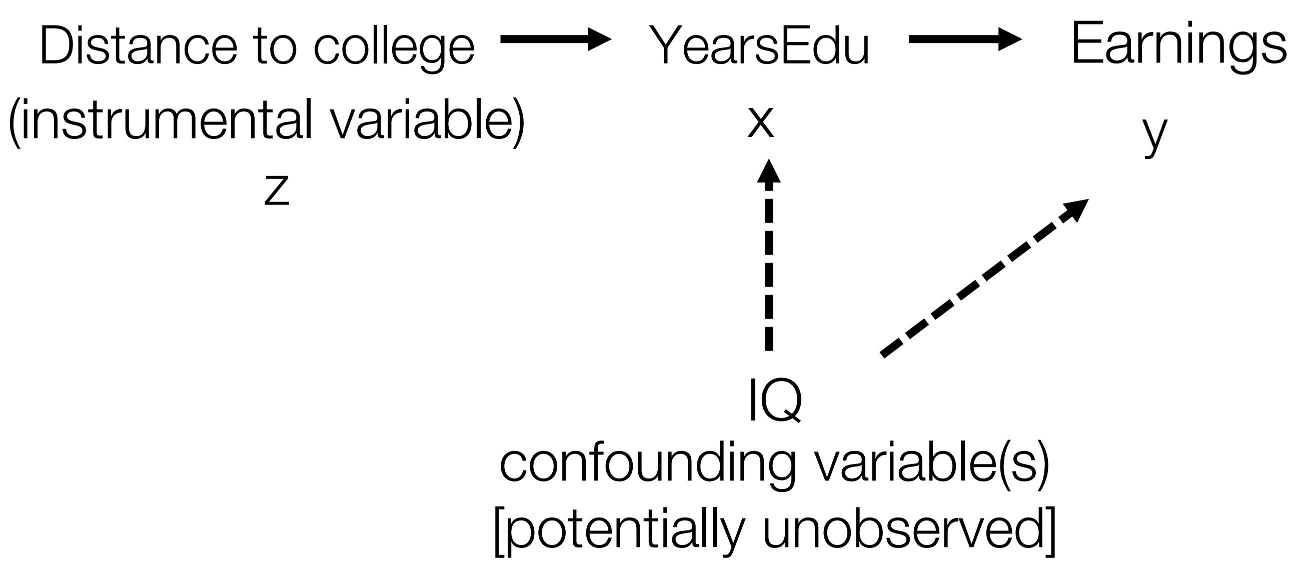


... natural randomizers

Images created with AI (Bing), April 14, 2024

# Instrumental variables

The method of instrumental variables (IV) is used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomized experiment. Intuitively, the method is used when an explanatory variable of interest is correlated with the error term, in which case ordinary least squares gives biased results. A **valid instrument (instrumental variable z) induces changes in the explanatory variable (x) but has no independent effect on the dependent variable (y),** allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.




**Estimation though two-stage least squares:**

Stage 1: generate predictions of YearsEdu:

YearsEdu_pred = $B_0$ + $B_1$ DisttoCollege + Error

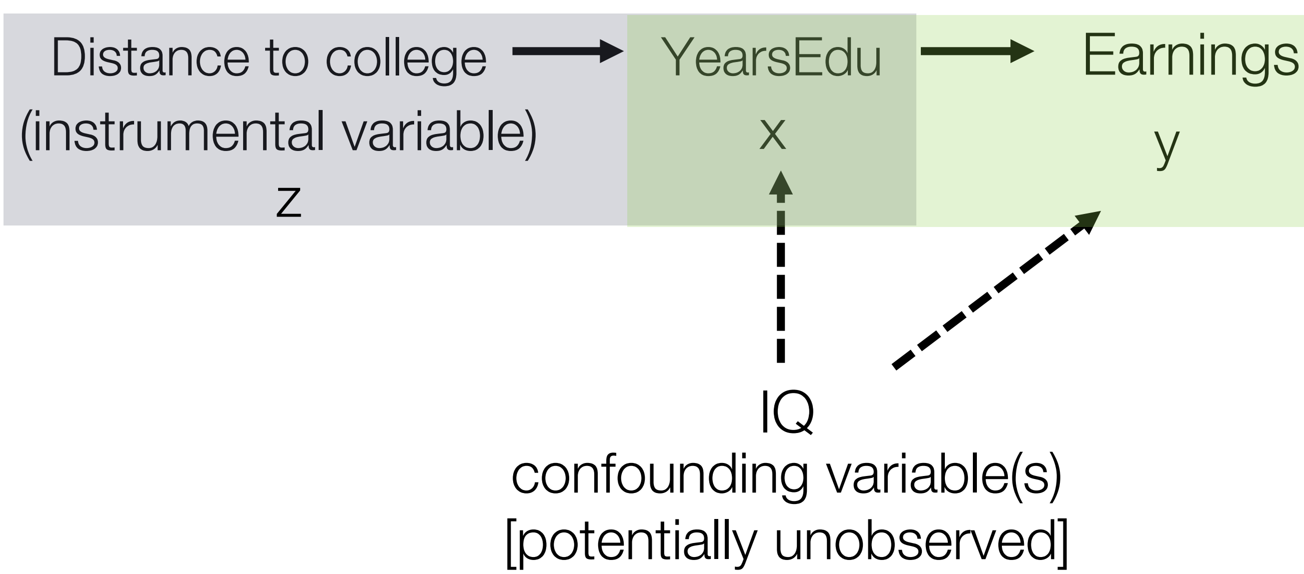
Stage 2: test whether YearsEdu_pred is significantly associated with earnings:

Earnings = $B_0$ + $B_1$ YearsEdu_pred + Error

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives, 15*(4), 69–85.

# Instrumental variables

The method of instrumental variables (IV) is used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomized experiment. Intuitively, the method is used when an explanatory variable of interest is correlated with the error term, in which case ordinary least squares gives biased results. A valid instrument (instrumental variable z) induces changes in the explanatory variable (x) but has no independent effect on the dependent variable (y), allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.




Estimation though two-stage least squares:

Stage 1: generate predictions of YearsEdu:

YearsEdu_pred = $B_0$ + $B_1$ Dist2College + Error

Stage 2: test whether YearsEdu_pred is significantly associated with earnings:

Earnings = $B_0$ + $B_1$ YearsEdu_pred + Error

**Problem**: Good instrumental variables (i.e., that are correlated with x but not any confounding variables) are hard to find, and you are only estimating a local average treatment effect (not an average treatment effect)…


Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives, 15*(4), 69–85.

# Instrumental variables

Introduction to Instrumental Variables (IV)


Econometrics - Instrumental Variables

# Instrumental variables

*Table 1*

**Examples of Studies That Use Instrumental Variables to Analyze Data From Natural and Randomized Experiments**

| *Outcome Variable* | *Endogenous Variable* | *Source of Instrumental Variable(s)* | *Reference* |
|---|---|---|---|
| | | *1. Natural Experiments* | |
| Labor supply | Disability insurance replacement rates | Region and time variation in benefit rules | Gruber (2000) |
| Labor supply | Fertility | Sibling-Sex composition | Angrist and Evans (1998) |
| Education, Labor supply | Out-of-wedlock fertility | Occurrence of twin births | Bronars and Grogger (1994) |
| Wages | Unemployment insurance tax rate | State laws | Anderson and Meyer (2000) |
| Earnings | Years of schooling | Region and time variation in school construction | Duflo (2001) |
| Earnings | Years of schooling | Proximity to college | Card (1995) |
| Earnings | Years of schooling | Quarter of birth | Angrist and Krueger (1991) |
| Earnings | Veteran status | Cohort dummies | Imbens and van der Klaauw (1995) |
| Earnings | Veteran status | Draft lottery number | Angrist (1990) |
| Achievement test scores | Class size | Discontinuities in class size due to maximum class-size rule | Angrist and Lavy (1999) |
| College enrollment | Financial aid | Discontinuities in financial aid formula | van der Klaauw (1996) |
| Health | Heart attack surgery | Proximity to cardiac care centers | McClellan, McNeil and Newhouse (1994) |
| Crime | Police | Electoral cycles | Levitt (1997) |
| Employment and Earnings | Length of prison sentence | Randomly assigned federal judges | Kling (1999) |
| Birth weight | Maternal smoking | State cigarette taxes | Evans and Ringel (1999) |

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives, 15*(4), 69–85.

# Natural experiments in economics and psychology

**Natural experiments:** “offer unique opportunities to combine features of randomized experiments and observational studies. A natural experiment is a “naturally” occurring event or condition (i.e., an event or condition not created by researchers) that affects some but not all units of a population. […] Natural experiments differ from (non-natural) randomized experiments in that participants are not randomly assigned to treatment and control groups by researchers, and researchers do not control the experimental manipulations and conditions.”

| Examples | Pros | Cons |
|---|---|---|
| • A longitudinal survey spans the occurrence of a relevant event (e.g., natural disaster) that affects parts of the population<br>• Educational reforms lead to the raising of minimum school-leaving age from one year to the next<br>• A lottery decides who gets drafted into military service during war times<br>• Lottery players who won large versus small sums of money | • An option when randomized experiments are unethical or unfeasible<br>• Limited demand effects<br>• More “realistic” effect sizes (cf. efficacy versus effectiveness)<br>• No self-selection bias (who volunteers for vaccine trials?) and lower participation burden | • Outcomes / dependent variables often not assessed immediately after the events/treatment occurred<br>• Hypotheses usually come after the event/treatment |

Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., ... & Rohrer, J. M. (2024). Natural experiments: Missed opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science, 7*(1), 1-15.

# Natural experiments in economics and psychology

**Table 1.** Review of a Random Sample of Economics and Psychology Articles

| | Economics | Psychology |
|---|---|---|
| Number of articles reviewed | 108 | 108 |
| Number of articles containing . . . | | |
| An empirical study | 88 | 96 |
| A randomized experiment | 17 | 42 |
| A natural experiment | 36[a] | 0 |
| A standard natural experiment with true randomization | 1 | 0 |
| A standard natural experiment with as-if randomization | 19 | 0 |
| An instrumental-variable design using a natural experiment with true randomization | 1 | 0 |
| An instrumental-variable design using a natural experiment with as-if randomization | 18 | 0 |
| A sharp regression-discontinuity design | 3 | 0 |
| A fuzzy regression-discontinuity design | 2 | 0 |

**What could explain the difference?**

"Reasons for the more frequent use of natural experiments in economics than psychology might be that randomized experiments are hardly feasible in macro-economics because researchers cannot experiment with countries' economies, rendering natural experiments such as public policies an attractive alternative to randomized experiments. Moreover, economists often use administrative observational data to exploit a natural experiment."

Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., ... & Rohrer, J. M. (2024). Natural experiments: Missed opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science, 7*(1), 1-15.

# A suitable natural experiment?

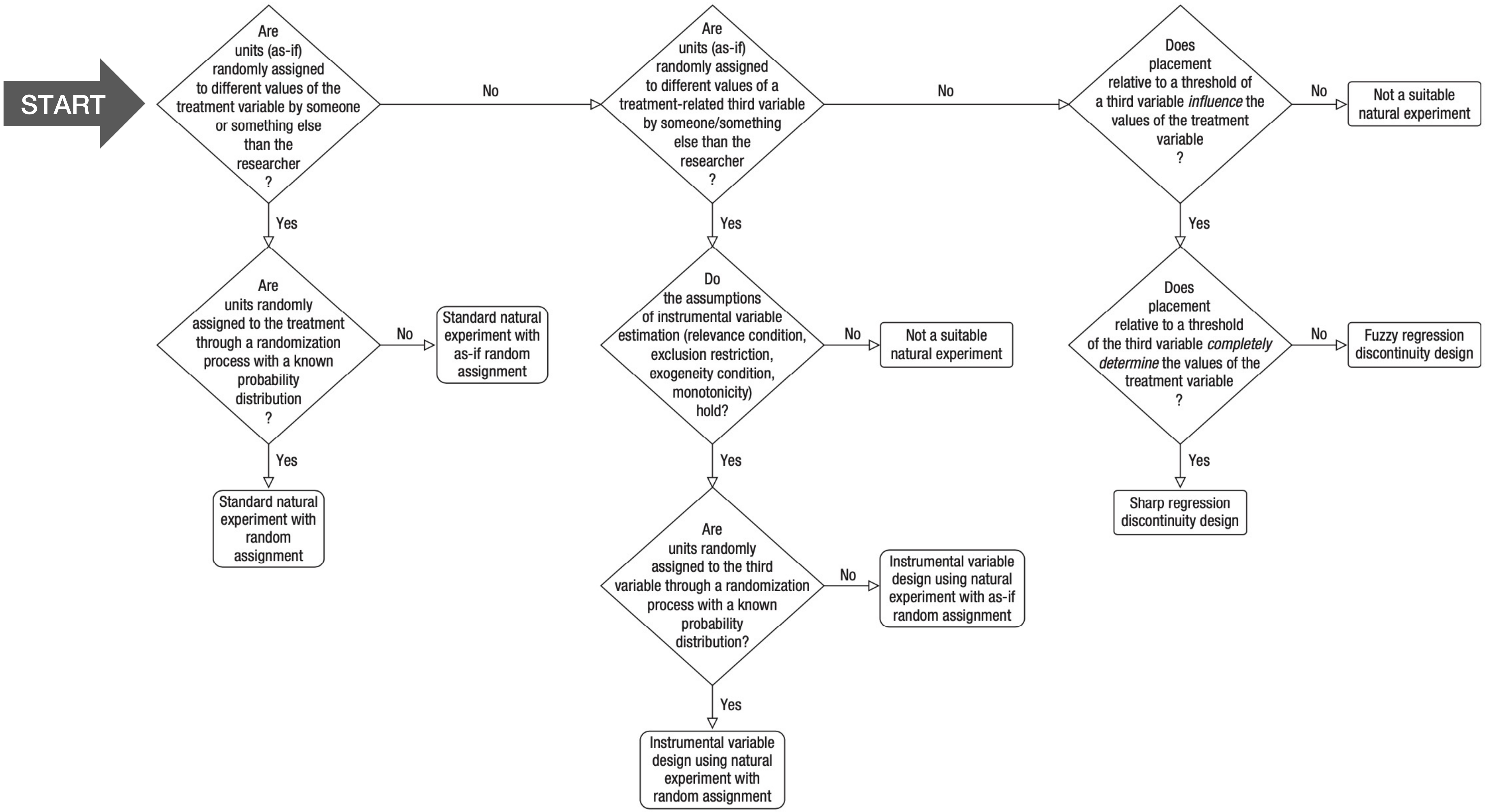
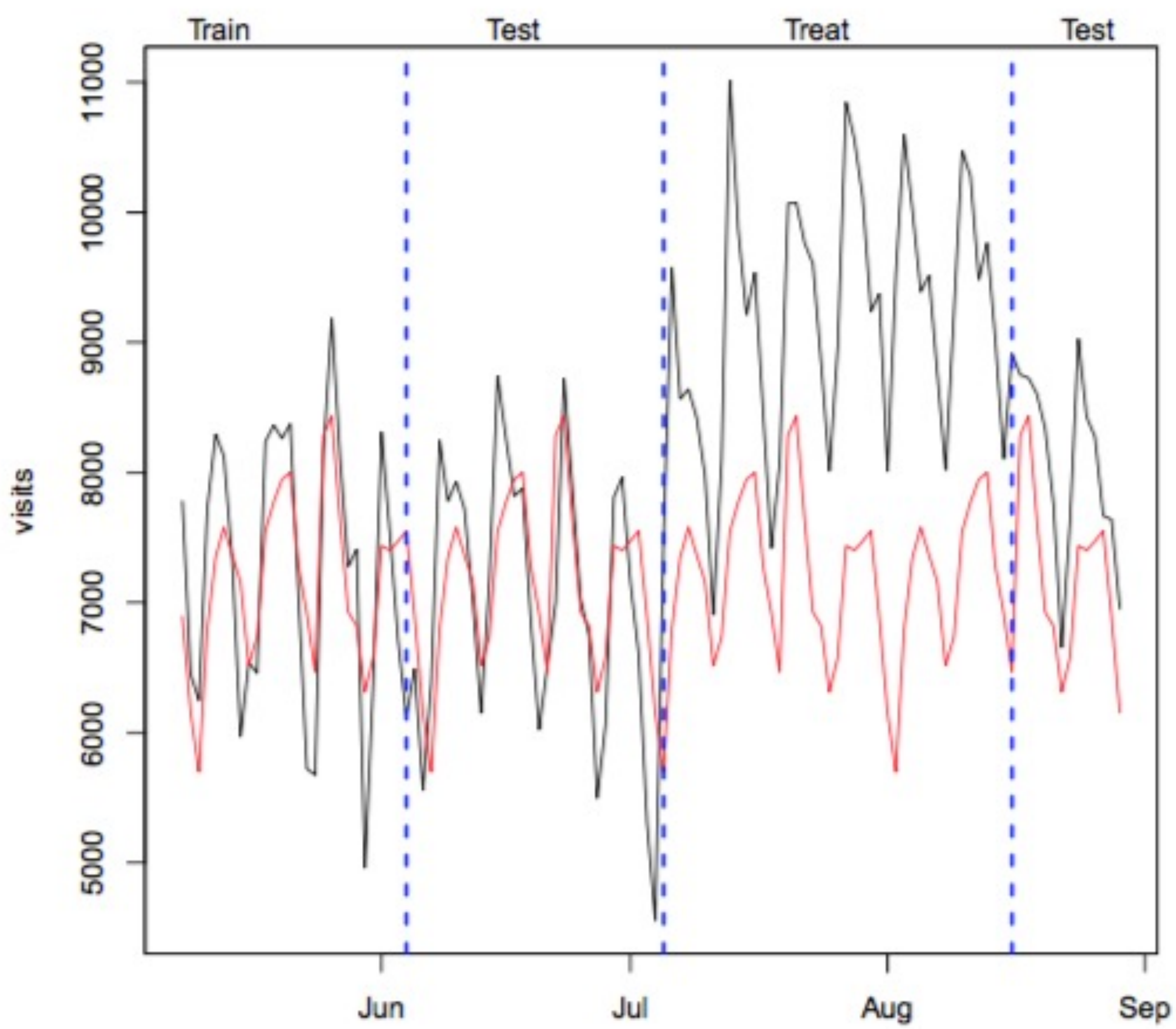



**Fig. 2.** Decision tree for identifying (types of) natural experiments.

Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., ... & Rohrer, J. M. (2024). Natural experiments: Missed opportunities for causal inference in psychology. *Advances in Methods and Practices in Psychological Science, 7*(1), 1-15.

# "Furious Six" statistical methods for causal inference?

## Using models as the control group (Train-test-treat-compare)




**An online advertiser might ask "if I increase my ad expenditure by some amount, how many extra sales do I generate?"**

A predictive statistical model (based on number of "searches" about topics related to the subject matter of the website) is estimated during the training period and its predictive performance is assessed during the test period. The extrapolation of the model during the treat period (red line) serves as a counterfactual. This counterfactual is compared with the actual outcome (black line), and the difference is the estimated treatment effect. When the treatment is ended, the outcome returns to something close to the original level.


Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

# Summary

- **Counterfactuals:** "The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment."

- **Causal Inference in Psychology:** There are many types of causal inference analyses that can be (and are) used in the behavioral sciences - in psychology, experiments and multiple regression from observational data are the most commonly used inference methods.

- **Furious Five / Six:** It is helpful to be aware of other methods (e.g., instrumental variables, regression discontinuity, difference in differences) and, more importantly, "the possibility of creatively utilizing the idiosyncratic features of any research situation in designing tests of causal hypotheses".

- **Natural Experiments:** Frequently exploited in economic research, natural experiments have untapped potential for psychological science (but also limitations!).