Making Policy with Data

An Introductory Course on Policy Evaluation

Lecture 3. Randomized Experiments

Instructor: Prof Yiqing Xu

April 18

Take-aways

1. Randomized Controlled Trials (**RCTs**) are the **gold standard** of causal inference

2. Random assignment eliminates selection bias

3. Successful implementation of an experiment is hard

Randomized Experiments Not New in Social Sciences

- Psychologists performing experiments in 1800s
- Political scientists Harold Gosnell used experiments to examine turnout in 1920s
  - Randomly assigned city blocks to receive mailed reminders
  - Turnout up 1% in the presidential elections of 1924, up 9% in the municipal election of 1925

Delayed Use Until Recently

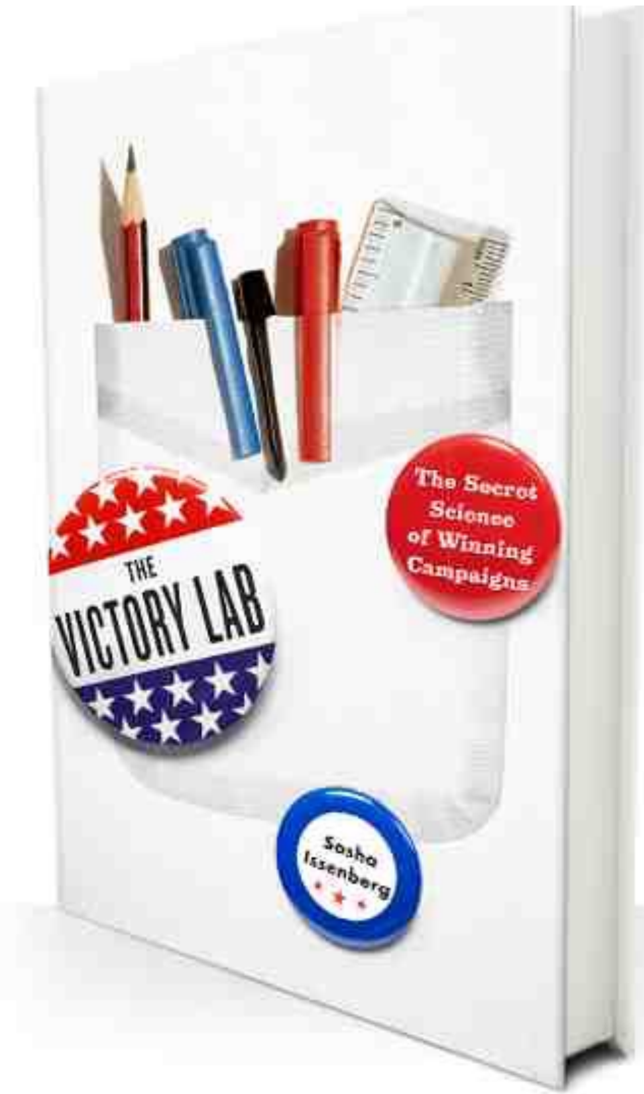
Traditionally and reasonably worries about artificiality

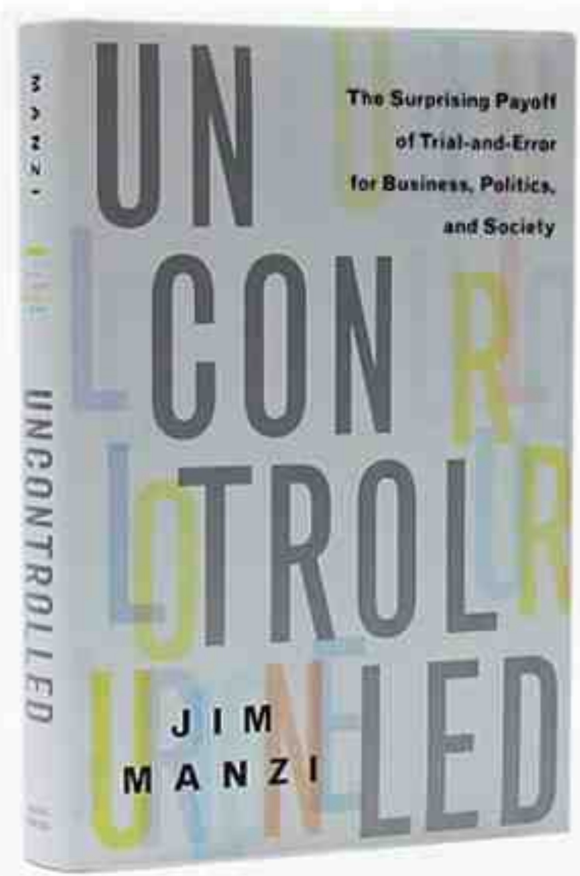
- Experimentation (by introducing artificiality) is suspect
- Control and manipulation not always possible for research

Experiments from Political Science and Economics

- Voter mobilization (Nickerson, Gerber and Green)
- Voting mechanisms (Olken)
- Health Insurance Reform (Finkelstein et al.)
- Race-based discrimination in labor markets (Bertrand and Mullainathan)
- Clientelistic vs Programmatic presidential campaigns (Wantchekon)
- Female Incumbents (Duflo)
- Information interventions for elites (Butler)
- Monitoring interventions (Ichino)
- Many more in the pipeline. . .

Experiments in Popular Culture

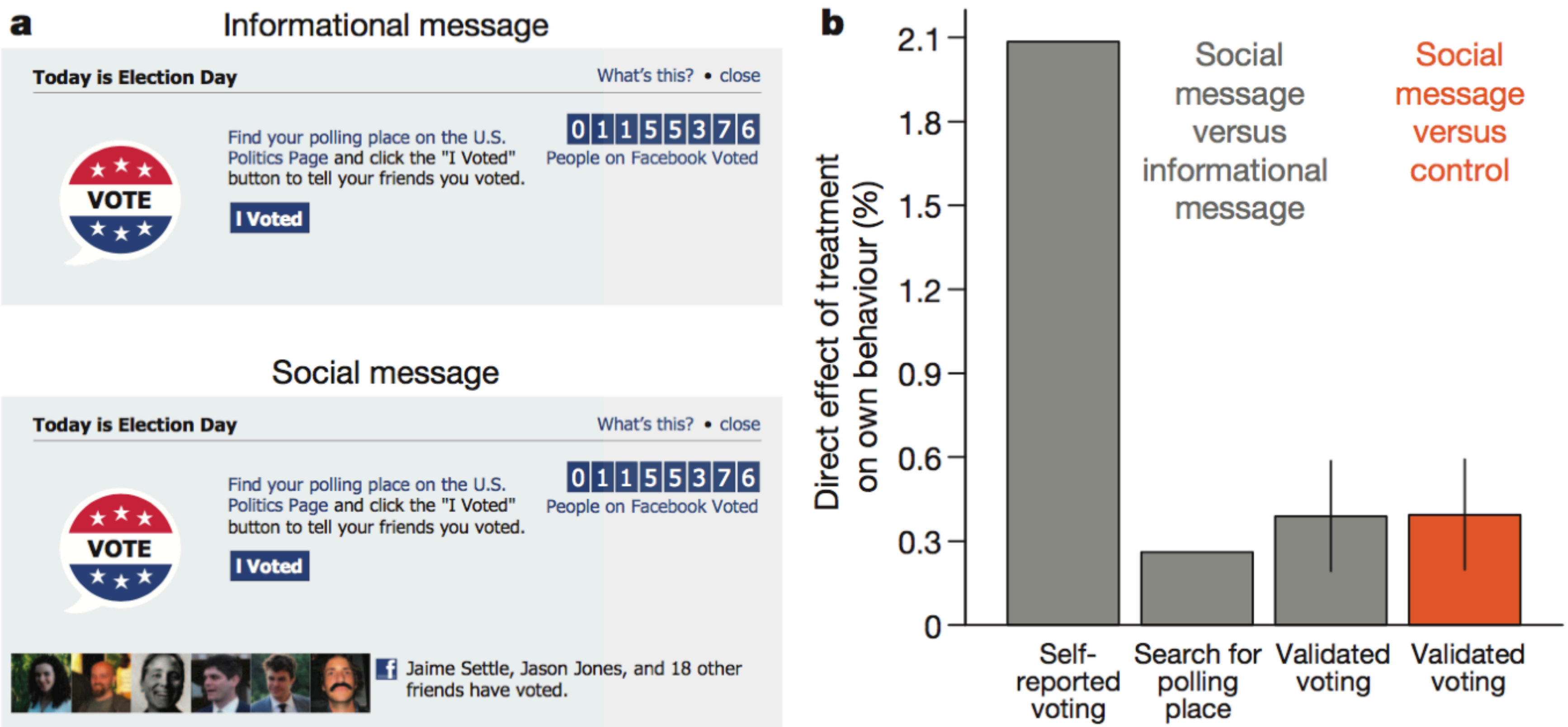




A Massive Social Pressure Experiment on Facebook

- More than 60 million people on Facebook saw a social, non-partisan “get out the vote” message at the top of their news feeds on Nov. 2, 2010.
- About 600,000 people, or one percent, were randomly assigned to see a modified “informational message,” identical in all respects to the social message except for pictures of friends.
- An additional 600,000 served as the control group and received no Election Day message from Facebook at all.

A Massive Social Pressure Experiment on Facebook



Agenda

- Identification and Estimation Under Random Assignment
- Experimental Design
- Inference Under Random Assignment

Identification Under Random Assignment

When selection bias goes away

Recall: Causation & Counterfactuals

- Key to successful program evaluation → estimate **counterfactual** by finding valid comparison groups
- Invalid comparison group → estimates of program effects mixed with estimates of other differences (selection bias)
- 2 methods particularly likely to give counterfeit counterfactual:
  - Comparing outcomes of participants ***before & after*** program
  - Compare outcomes of those ***with & without*** program
- By contrast, randomization is the **gold standard** of impact evaluation
- Random assignment provides robust estimate of counterfactual

Recall: Potential Outcomes Framework

- Outcome

  $Y_i =$ Observed outcome for unit i

- Treatment

  D_i: Indicator of whether unit i received treatment

$$D_i = \begin{cases} 1 & \text{unit } i \text{ received treatment} \\ 0 & \text{unit } i \text{ did not receive treatment} \end{cases}$$

- Potential Outcomes

  Y_{1i} : Potential outcome for i with treatment

  Y_{0i} : Potential outcome for i without treatment

Selection Bias

- Recall the selection problem when comparing the mean outcomes for the treated and the untreated:

$$
\begin{aligned}
E[Y_i|D=1] - E[Y_i|D_i=0] \\
&= E[Y_{1i}|D_i=1] - E[Y_{0i}|D_i=0] \\
&= \underbrace{E[Y_{1i} - Y_{0i}|D_i=1]}_{\text{ATT}} + \underbrace{\{E[Y_{i0}|D_i=1] - E[Y_{0i}|D_i=0]\}}_{\text{BIAS}}
\end{aligned}
$$

- As a result of randomization, the selection bias term will be 0.

- The treatment and control group will tend to be similar along all characteristics (identical in expectation), including the potential outcomes under the control condition

Identification Under Random Assignment

Before treatment is assigned:

i	$\Pr(D_i = 1)$	D_i	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$
1	$\frac{50}{100}$	?	10	4	6
2	$\frac{50}{100}$	?	1	2	-1
$\vdots$	$\frac{50}{100}$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
99	$\frac{50}{100}$	?	3	3	0
100	$\frac{50}{100}$	?	5	2	3

After treatment is assigned:

i	$\Pr(D_i = 1)$	D_i	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$
1	$\frac{50}{100}$	0	?	4	?
2	$\frac{50}{100}$	1	1	?	?
$\vdots$	$\frac{50}{100}$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
99	$\frac{50}{100}$	1	3	?	?
100	$\frac{50}{100}$	0	?	2	?

Identification Under Random Assignment

Identification Assumption

$(Y_1, Y_0) \perp\!\!\!\perp D$ *(random assignment)*

Identfication Result

Problem: $\alpha_{ATE} = \mathbf{E}[Y_1 - Y_0]$ *is unobserved. But given random assignment*

$$\begin{aligned} \mathbf{E}[Y|D=1] &= \mathbf{E}[D \cdot Y_1 + (1-D) \cdot Y_0 | D=1] \\ &= \mathbf{E}[Y_1|D=1] \\ &= \mathbf{E}[Y_1] \end{aligned}$$

$$\begin{aligned} \mathbf{E}[Y|D=0] &= \mathbf{E}[D \cdot Y_1 + (1-D) \cdot Y_0 | D=0] \\ &= \mathbf{E}[Y_0|D=0] \\ &= \mathbf{E}[Y_0] \end{aligned}$$

$$\alpha_{ATE} = \mathbf{E}[Y_1 - Y_0] = \mathbf{E}[Y_1] - \mathbf{E}[Y_0] = \underbrace{\mathbf{E}[Y|D=1] - \mathbf{E}[Y|D=0]}_{\textit{Difference in Means}}$$

Identification Under Random Assignment

i	Y_{1i}	Y_{0i}	Y_i	D_i
1 2	$E[Y_1 \vert D=1]=1$	$E[Y_0 \vert D=1]=?$	2 0	1 1
3 4	$E[Y_1 \vert D=0]=?$	$E[Y_0 \vert D=0]=2$	1 3	0 0

$$\alpha_{ATT} = E[Y_1 - Y_0 | D = 1]$$

$(Y_1, Y_0) \perp\!\!\!\perp D$ implies $E[Y_0|D=1] = E[Y_0|D=0]$

$$\begin{aligned}
\alpha_{ATT} &= E[Y_1 - Y_0|D=1] \\
&= E[Y_1|D=1] - E[Y_0|D=1] \\
&= E[Y_1|D=1] - E[Y_0|D=0] \\
&= E[Y|D=1] - E[Y|D=0]
\end{aligned}$$

Identification Under Random Assignment

i	Y_{1i}	Y_{0i}	Y_i	D_i
1 2	$E[Y_1 \mid D=1]=1$	$E[Y_0 \mid D=1]=?$	2 0	1 1
3 4	$E[Y_1 \mid D=0]=?$	$E[Y_0 \mid D=0]=2$	1 3	0 0

$$\alpha_{ATT} = E[Y_1 - Y_0 \mid D=1]$$

$(Y_1, Y_0) \perp\!\!\!\perp D$ implies $E[Y_0 \mid D=1] = E[Y_0 \mid D=0]$

$$\begin{aligned}
\alpha_{ATT} &= E[Y_1 - Y_0 \mid D=1] = E[Y_1 \mid D=1] - E[Y_0 \mid D=0] \\
&= E[Y_1] - E[Y_0] = E[Y_1 - Y_0] \\
&= \alpha_{ATE}
\end{aligned}$$

Benefits of Random Assignment

- Budget / capacity constraints, so often don't fully reach intended population

- Random assignment provides valid comparison group
  - Ration by chance, rather than observables or first-come, first-serve
  - Random assignment yields two groups with high probability of being statistically identical, if sufficient N
  - If N large, random assignment yields statistically equivalent averages for observables **AND** unobservables

Estimation Under Random Assignment

Consider a randomized trial with N individuals.

Estimand

$\alpha_0 = E[Y_1 - Y_0] = \alpha_{ATE} = \alpha_{ATT} = \mathbf{E}[Y|D=1] - \mathbf{E}[Y|D=0]$

Estimator

By the analogy principle we use $\widehat{\alpha} = \bar{Y}_1 - \bar{Y}_0$ *where*

$$\bar{Y}_1 = \frac{\sum Y_i \cdot D_i}{\sum D_i} = \frac{1}{N_1} \sum_{D_i=1} Y_i \,;$$

$$\bar{Y}_0 = \frac{\sum Y_i \cdot (1 - D_i)}{\sum (1 - D_i)} = \frac{1}{N_0} \sum_{D_i=0} Y_i$$

with $N_1 = \sum_i D_i$ *and* $N_0 = N - N_1$.

Randomized Experiments and Regression

- Randomized experiments can be analyzed using regression, though when covariates are added there are some subtleties about interpretation.

$$
\begin{aligned}
Y_i &= \alpha + bD_i + \epsilon_i \\
Y_i &= \underbrace{\bar{Y}_0}_{\alpha} + \underbrace{(\bar{Y}_1 - \bar{Y}_0)}_{b} D_i + \underbrace{\{(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)]\}}_{\epsilon} \\
&= D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}
\end{aligned}
$$

Adjusting Covariates

- Practitioners often run some variant of the following model with experimental data:

$$Y_i = \alpha + bD_i + X_i\beta + \epsilon_i$$

- Why include the X_i when experiments "control" for covariates by design?
  - Correct for chance **covariate imbalances** (bad luck)
  - **Increase precision:** reduce variation in the outcome accounted for by pre-treatment characteristics, thus making it easier to attribute remaining differences to the treatment.
- Generally, ATE estimates are robust to model specification. Never control for **post-treatment covariates**!

Adjusting Covariates

- Lin (2013) discusses covariate adjustment via a regression of the following form:

$$y_i = \alpha + bD_i + \beta_1 \cdot (x_i - \bar{x}) + \beta_2 \cdot D_i \cdot (x_i - \bar{x}) + \epsilon_i$$

- **Consistent** estimator for ATE
  - Getting closer and closer to ATE as the sample becomes larger.

- Cannot hurt precision and if covariates are predictive of the outcomes, then will likely **increase precision**.
  - Gives you smaller standard errors for the ATE estimates

Experimental Design

Getting into the Weeds

The No Interference Assumption

Assumption

Observed outcomes are realized as

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}$$

- Embedded in this formulation is the assumption that potential outcomes for unit i are unaffected by treatment assignment for unit j.
- Assumption known as **Stable Unit Treatment Value Assumption (SUTVA)**
- Examples: vaccination, fertilizer on plot yield, communication

SUTVA

SUTVA has two parts

1. No interference

Units do not interfere with each other: treatment applied to one unit does not effect the outcome for another unit.

2. Only one version of each treatment level exists

Potential outcomes is well defined.

Unit of Analysis

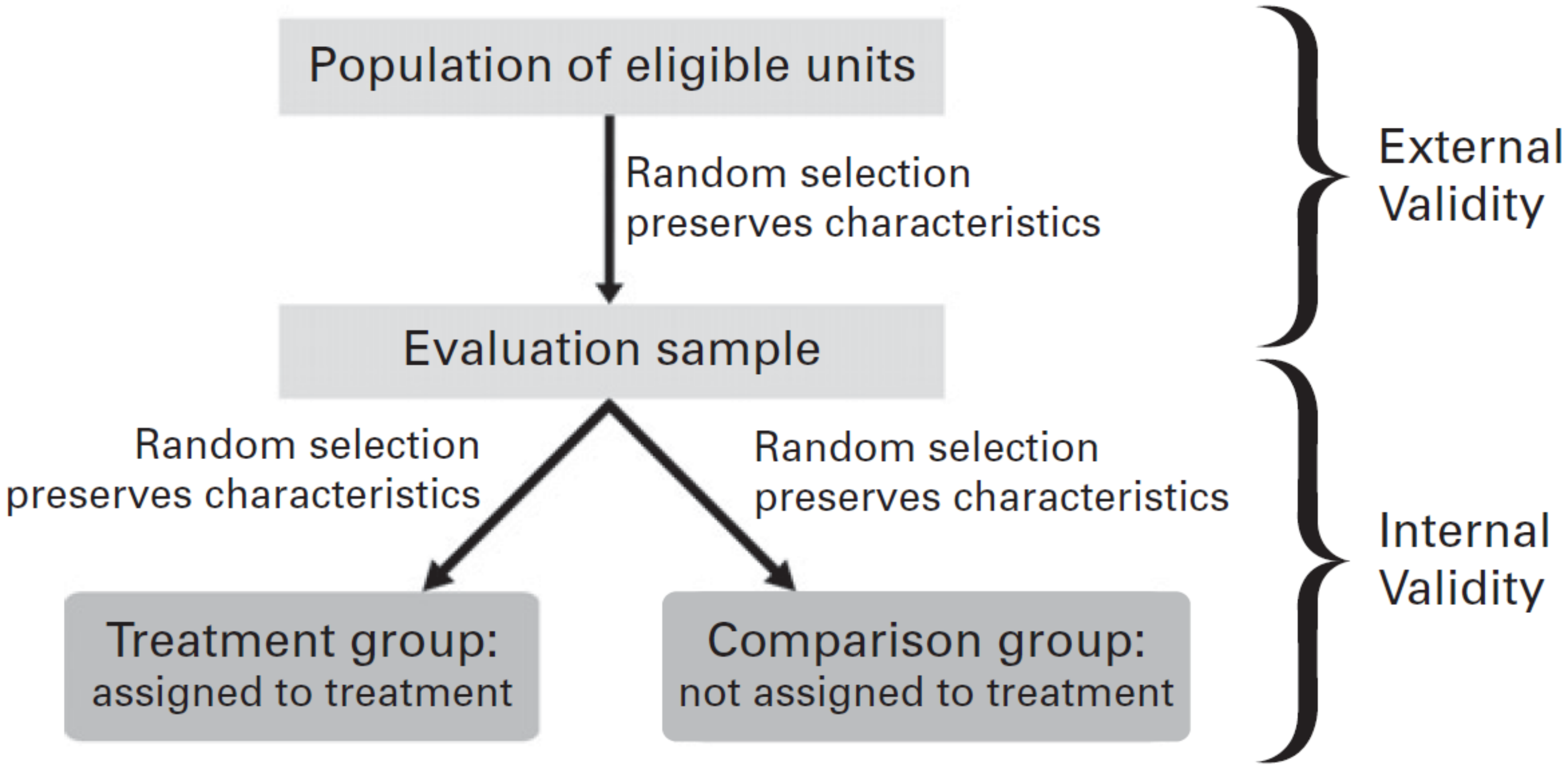
- Unit of analysis and unit of randomization (individuals, groups, institutions, etc)?
  - Choice of analytic level determines what the study has the capacity to demonstrate.
  - Example: randomize school vouchers at the level of the individual or at the level of the community? Do want to know how students respond to new environment or or how schools respond to competition?
  - Can also help with SUTVA (e.g. interactions within and between schools)
- How many treatments?
- How many units?
- How many treated and how many controls?
- Is background information available? If so, how can it be used?

Internal & External Validity

- ***External validity***: evaluation sample accurately represents population of eligible units
    - Random sampling of population, so evaluation sample representative of population
- ***Internal validity***: valid comparison group used, so no confounding factors in estimated impact
    - Random assignment such that comparison group statistically equivalent to T at baseline

Source: Gertler, 2011.

Internal & External Validity



Source: Gertler, 2011.

Partners for Experiments

- Randomization almost always requires working with a partner
- ***Governments*** intend to serve all of eligible population, but may experiment with pilot before scaling up
  - Requires high-level consensus
  - May face difficulties from officials with upset constituents
  - Wider geographic scope, more likely results will influence policy
- ***NGOs*** less subject to discrimination problems, as not typically meant to serve entire population
  - More flexible, so can monitor implementation and affect design
  - Often partner of choice, as more willing
  - But are results dependent on organizational culture?
- ***For profit firms***, especially in the world of micro-credit

Ethics

- **Respect for persons:** Participants in most circumstances must give informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules can be waived.
- **Benevolence:** Avoid knowingly doing harm. Does not mean that all risk can be eliminated, but possible risks must be balanced against overall benefits to society of the research.
  - Note that the existence of a control group might be construed as denying access to some benefit.
  - But without a control group, generating reliable knowledge about the efficacy of the intervention may be impossible.
- **Justice:** Important to avoid situations where one group disproportionately bears the risks and another stands to received all the benefits.
  - Evaluate interventions that are relevant to the subject population

Source: Gertler, 2011.

Ethics

- IRB approval is required in almost all circumstances.
- If running an experiment in another country, you need to follow the local regulations on experimental research.
    - Often poorly adapted to social science.
    - Or legally murky whether or not approval is required.
- Still many unanswered questions and lack of consensus on the ethics of field experimentation within Political Science!
    - Be prepared to confront wildly varying opinions on these issues.

Source: Gertler, 2011.

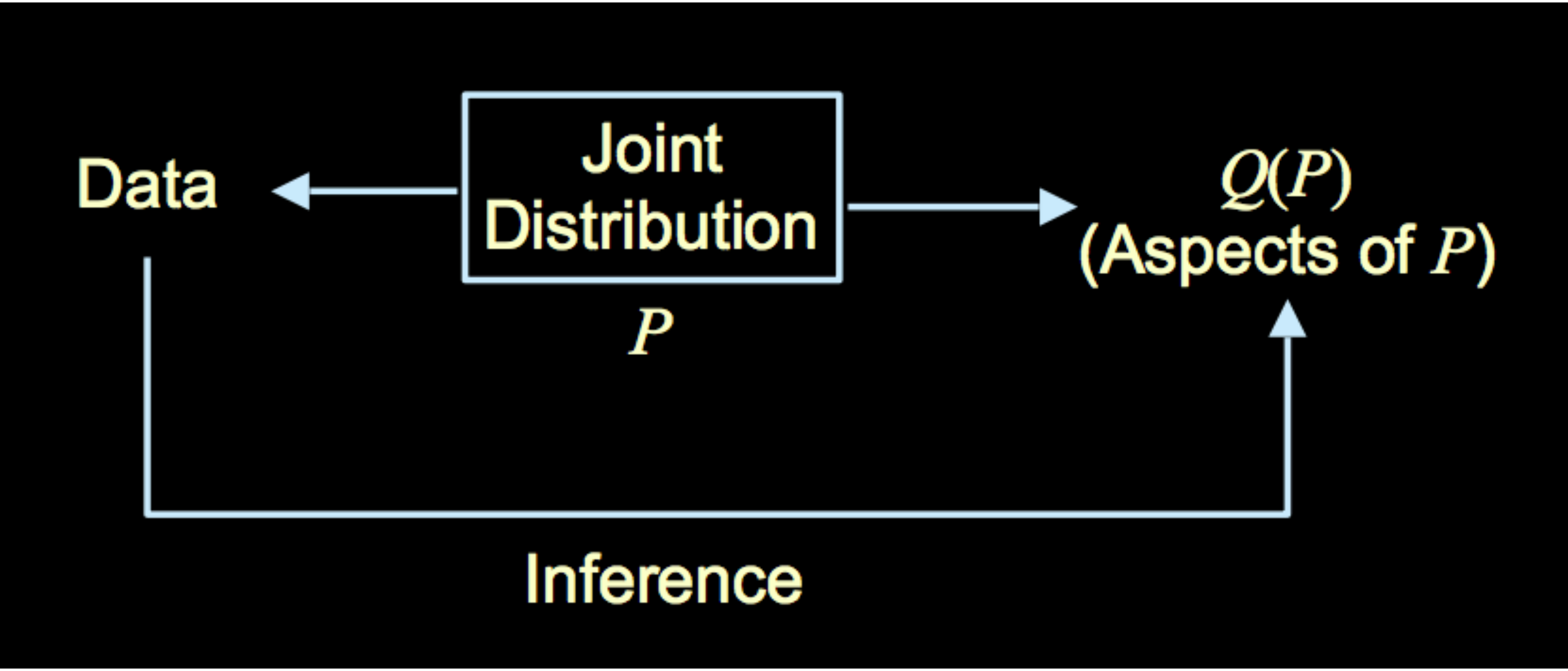
Inference Under Random Assignment

What can we make of experimental results?

Estimands, Estimators, and Estimates

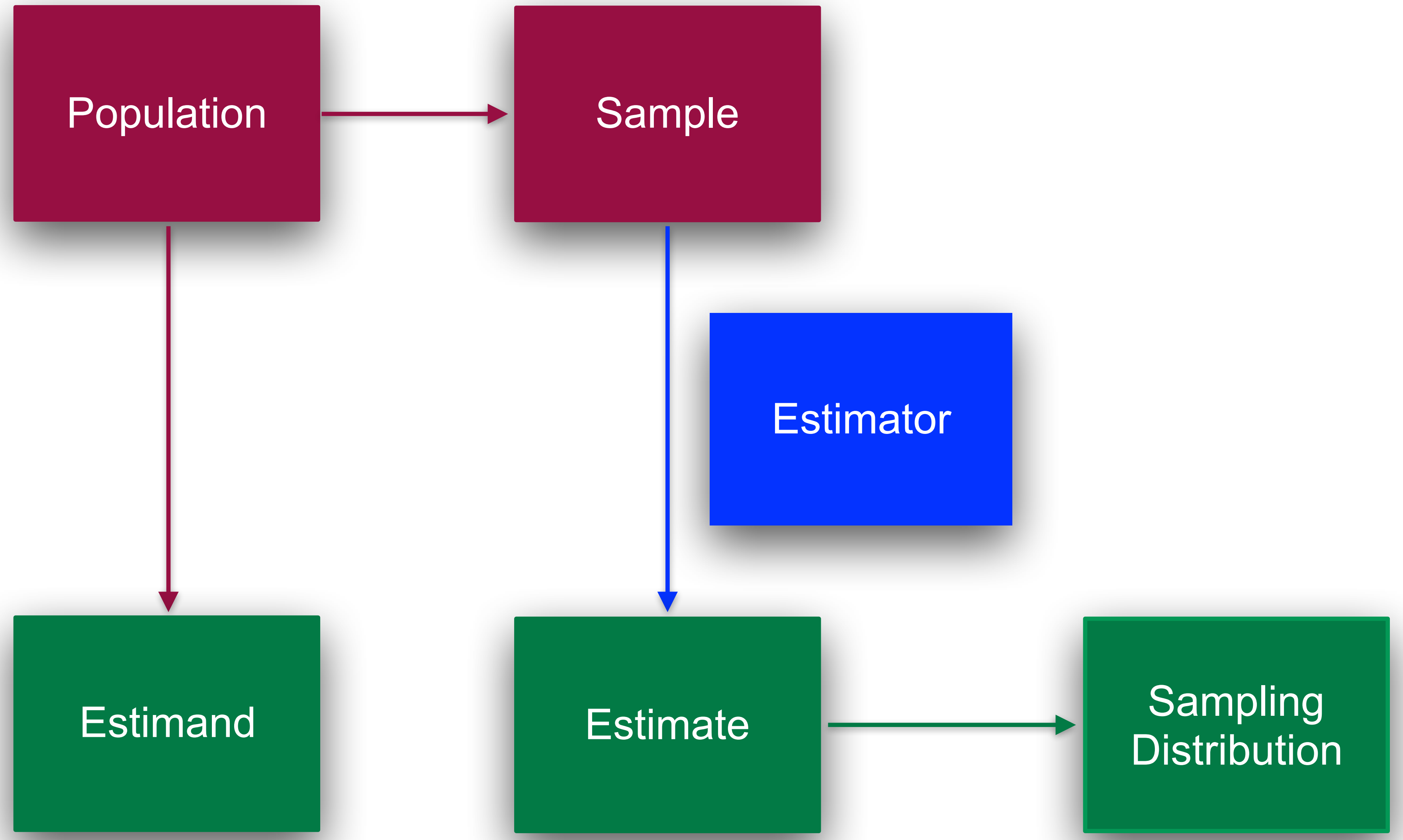
- We refer to the (unobserved) population characteristic that we aim to learn about as the **estimand** T
    - Example: ATE
- To learn about estimands, we use functions of the sample data called **estimators**
    - Example: Difference-in-Means estimator
- The values taken by the estimators for particular samples are called **estimates**
    - Example: Difference in means of data from an experiment

Statistical Inference

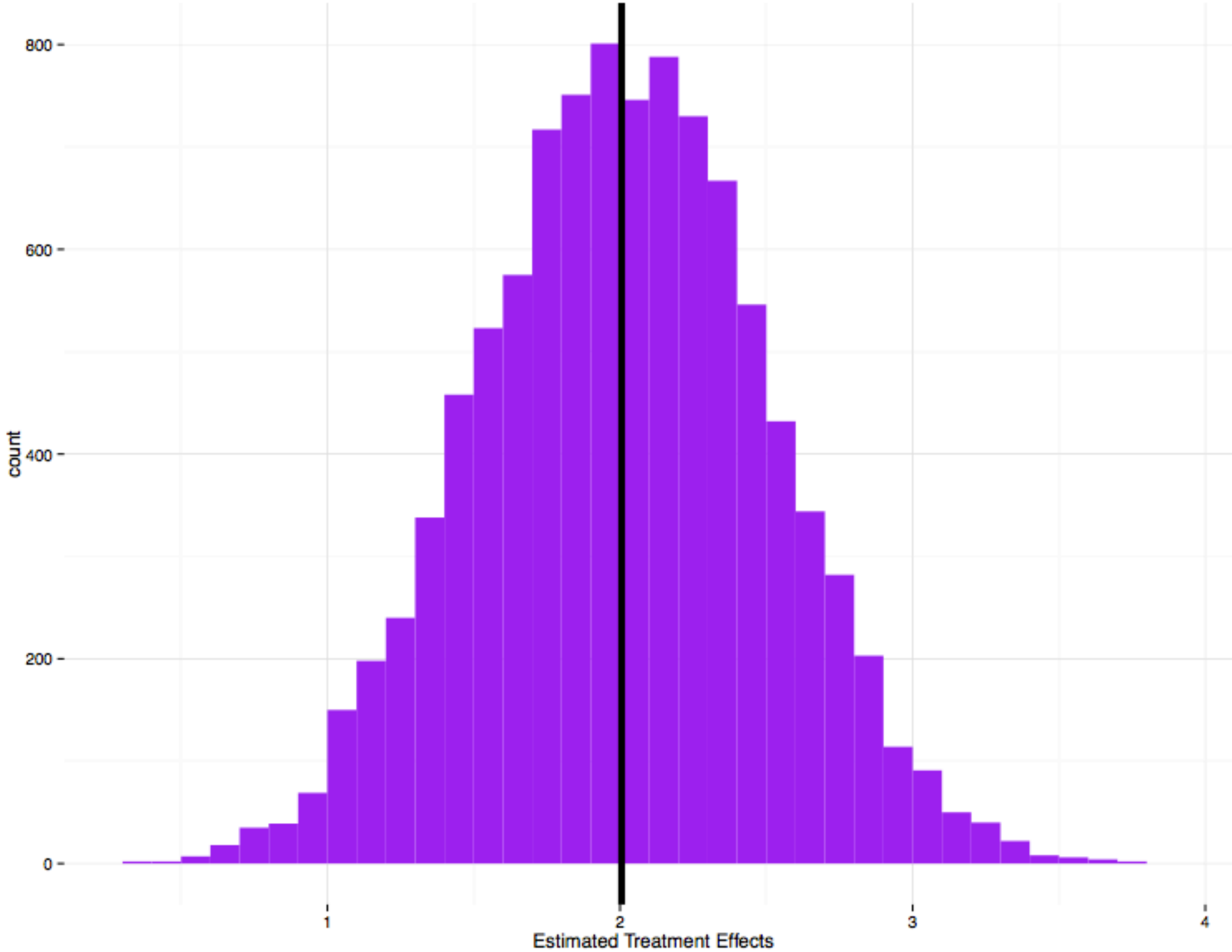


Elias Bareinboim

Internal & External Validity



Sampling Distribution



Sampling Distribution

- In order to assess the properties of an estimator, we assume it has a distribution under “repeated sampling”, and we call this distribution a **sampling distribution**

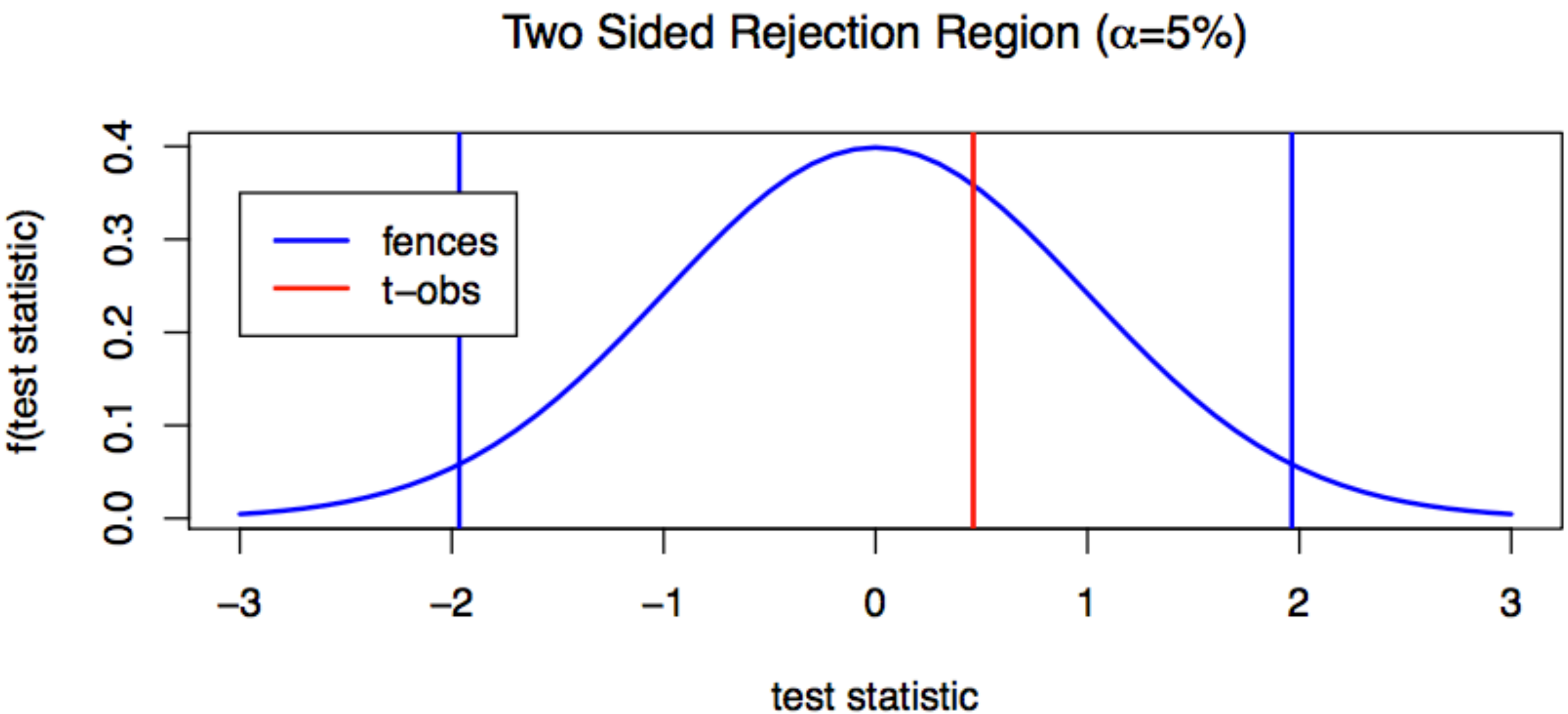
- We use information from one sample to approximate the unobserved sampling distribution

- Then we use both (1) the estimate and (2) the approximate sampling distribution under the Null hypothesis to conduct **hypothesis testing**

Hypothesis Testing Procedure

- Specify a null hypothesis, e.g. ATE = 0
- Pick a test level, e.g. 5%
- Choose a test statistic, e.g. *t*-statistic
- Derive the null distribution (using your data)
- Using the null distribution, look up the critical value that corresponds to your chosen test level and your test statistic (based on your data)

Internal & External Validity

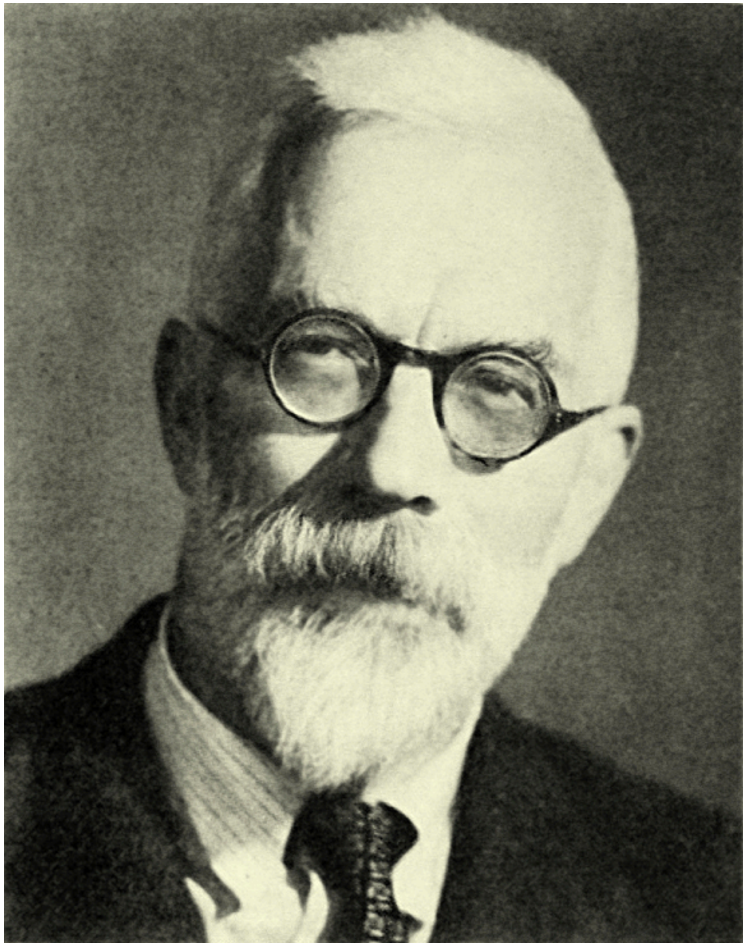


"The treatment effect is significant at the 5% level" means that, imagine we conduct the same experiment again and again, with the current ATE estimate, if we reject the NULL hypothesis that the ATE is 0, we will be making mistakes less than 5 times out of 100 trials.

Summary

- Random assignment solves the **identification problem** for causal inference based on minimal that we can even control as researchers assumptions
- Random assignment balances **observed and unobserved** confounders, which is why it is considered the gold standard for causal inference
- Statistical analysis is almost trivial and results are usually not model dependent, since confounders are controlled for "by design"
- Design features can help to improve inferences
- Always important to think about theory and external validity prior to experimentation.

Fisher and Smoking



The father of RCTs:

“Smoking and lung cancer share a common genetic origin.” —It is selection bias.

This is wrong.