Causal Inference

Lecture #1

Juraj Medzihorsky

SGIA & RMC
Durham University

9 January 2023

Drawing conclusions about causes and effects.

Looking and acting

Looking: Why? How?

Acting: What if? What to do?

Goals

- To know, and be able to use most of the standard toolbox for causal inference.
- To know where to look for more advanced things.
- To spot common errors and deficiencies in applied causal inference.
- To navigate

  the trade-offs inescapable in applied causal inference

  and common evidence hierarchies

  to maximize insight and impact

Goals: seeing behind the salesmanship and future-proofing

An Alchemist's Laboratory. 18^{th} century. Follower of David Teniers II.

Source: `https://sciencehistory.org/stories/magazine/the-secrets-of-alchemy/`

Structure

Meeting			Main content
1.	9 Jan	Lect #1	Introduction, refresh
2.	16 Jan	Sem #1	Randomization, matching
3.	23 Jan	Lect #2	Regression, diff-in-diff
4.	30 Jan	Sem #2	Regression, DiD, NOC
5.	6 Feb	Lect #3	Synthetic control method
6.	13 Feb	Sem #3	Synthetic control method
7.	20 Feb	Lect #4	Regression discontinuity design
8.	27 Feb	Sem #4	Regression discontinuity design
9.	5 Mar	Lect #5	Instrumental variables
10.	12 Mar	Sem #5	Instrumental variables, review

Structure

Lectures

- Readings
- Slides

Seminars

- Annotated R code, live coding
- You can follow on your own machines or just watch
- Code can include take-home exercises
- Also some readings and slides

You can use any software you want for the assessments.

Readings: The textbooks

Readings: Additional readings

- Articles
  - Focus on data + methods
  - Checking replication archives can help

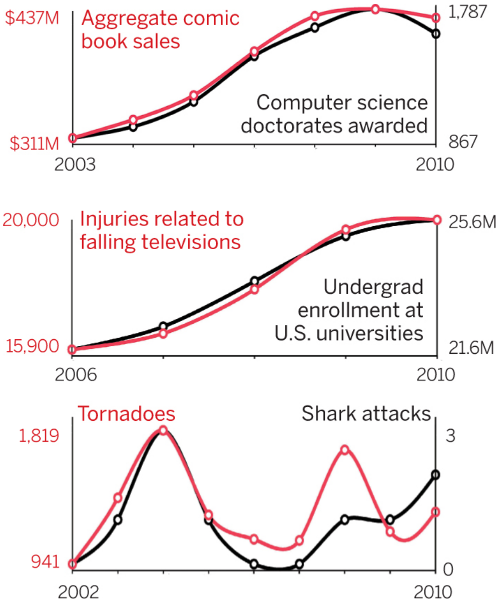
Assessments

Formative

- Replication of a causal analysis from a published article
- Keeping looking for recent articles on your favorite topics
- 1,500 words
- 6 March 2024, noon

Summative (100%)

- Report on an original causal analysis
- Never too early to start looking for data
- 3,000 words
- 26 April 2024, noon



https://doi.org/10.1126/science.aac5518

Conclusions = Assumptions + Data

Estimate = Truth + Noise + Bias

$$
\begin{aligned}
\textit{Conclusion} &= \textit{Estimate} \\
\textit{Assumptions} + \textit{Data} &= \textit{Truth} + \textit{Noise} + \textit{Bias}
\end{aligned}
$$

Detectives, Farmers, and Causal Questions

🕵🏿‍♂️ Detective	👩🏻‍🌾 Farmer
"Who's done it?"	*"Will the fertilizer help?"*
Responsibility	Intervention
What caused y_i?	Does X affect Y?
Backwards	Forwards
Causes of effects	Effects of causes
Most humans	R. A. Fisher

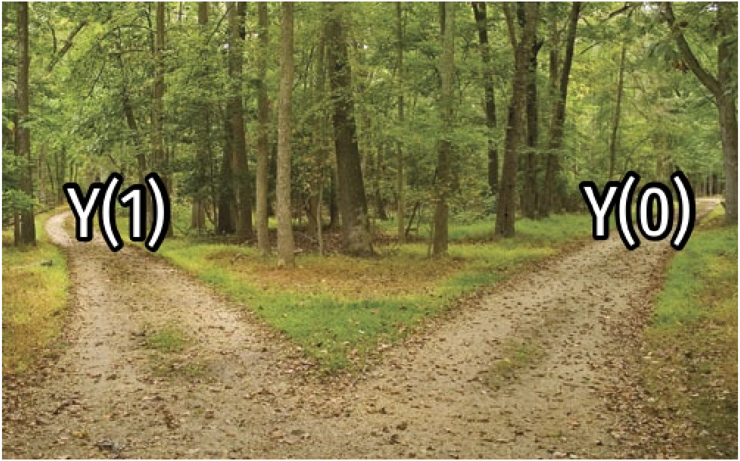
Looking back and looking forth

Looking back

"O.K., I admit it, we're lost, but the important thing is to remain focussed on whose fault it is."

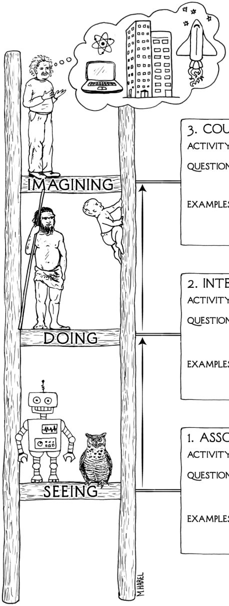
Source: *The New Yorker*

Looking forth



Source: Matthew Blackwell

Pearl's Ladder of Causality



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done …? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see …?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

Source: *The Book of Why* by Pearl and Mackenzie

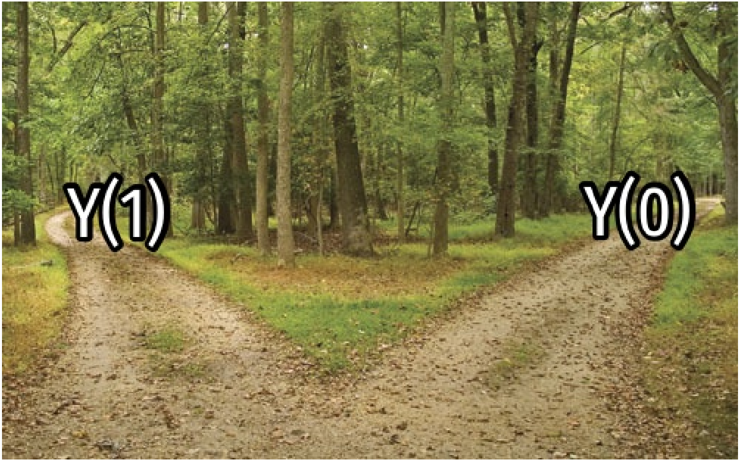
Contemporary approaches to causal inference

- Potential Outcomes Framework, a.k.a. Neyman-Rubin-(Holland)
- Structural Causal Model (Pearl et al.)
- Decision-theoretic approach of Philip Dawid
- Naive regressionism-controllism
- ...

Contemporary approaches to causal inference

- Potential Outcomes Framework, a.k.a. Neyman-Rubin-(Holland)
- Structural Causal Model (Pearl et al.)
- Decision-theoretic approach of Philip Dawid
- Naive regressionism-controllism
- ...

Causation and potential outcomes



Source: Matthew Blackwell

The Potential Outcomes Framework

A general statistical framework for the analysis of cause and effect

- Based on ideas older than the field of statistics
- Jerzy Neyman (1923) introduced the potential outcomes notation for experiments
- R.A. Fisher (1925) proposed actually randomising treatments to units
- Donald Rubin (1974) extended the potential outcomes framework [a.k.a. the “Rubin Causal Model” (Holland 1986)] to observational studies

The Potential Outcomes Framework

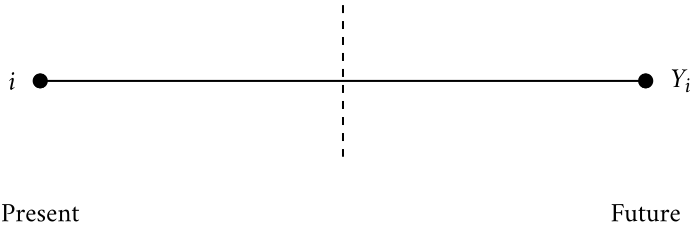
Potential Outcomes $\longrightarrow$ Potential Future Worlds

The Potential Outcomes Framework

Potential Outcomes $\longrightarrow$ Potential Future Worlds

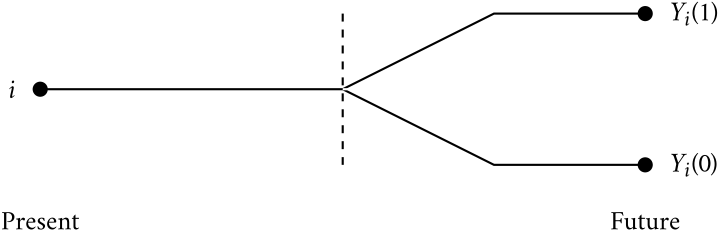
The Potential Outcomes Framework

Potential Outcomes $\longrightarrow$ Potential Future Worlds



The Potential Outcomes Framework

Potential Outcomes $\longrightarrow$ Potential Future Worlds



Causation and manipulability

- If unit i receives $d \in \mathcal{D}$, then $Y_i = Y_i(D_i = d) = Y_i(d)$
- *"No causation without manipulation"* Holland (1986)

  D_i must be manipulable at least in principle:

  it can be changed without immediately changing other relevant attributes of unit i.

Types of Effects

Average Treatment Effect

$$ATE = \mathbb{E}\left[Y_i(1) - Y_i(0)\right]$$

Average Treatment Effect on the Treated

$$ATT_1 = \mathbb{E}\left[Y_i(1) - Y_i(0) | D_i = 1\right]$$

Conditional Average Treatment Effect

$$CATE = \mathbb{E}\left[Y_i(1) - Y_i(0) | \boldsymbol{X}_i = \boldsymbol{x}\right]$$

Quantile Treatment Effect

$$QTE = \mathbb{Q}_q\left[Y_i(1)\right] - \mathbb{Q}_q\left[Y_i(1)\right]$$

and many others ...

Key assumptions for defining effects

1. Causal ordering

$$D_i \longrightarrow Y_i$$

2. Consistency

$$Y_i = Y_i(d) \text{ if } D_i = d$$

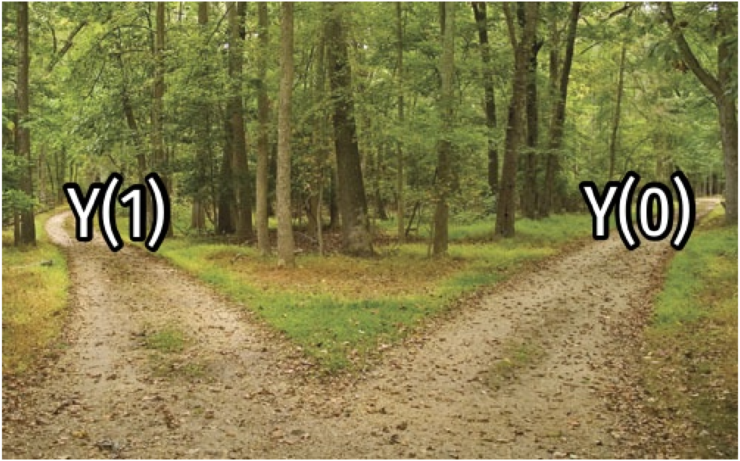
3. No interference

$$Y_i(D_1, D_2, \ldots, D_n) = Y_i(D_i)$$

Stable Unit Treatment Value Assumption (SUTVA): 2. & 3. combined.

Relaxing 3. a highly active research area.

Fundamental Problem of Causal Inference



Source: Matthew Blackwell

Example: World Bank Poverty Reduction Program

- Future GDP Per Capita

Unit	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
Denmark	\$41,932	\$43,445	\$1,513
Liberia	\$806	\$878	\$72
Australia	\$42,063	\$43,544	\$1,481
Afghanistan	\$988	\$1,946	\$958
South Korea	\$31,079	\$33,140	\$2,061
Haiti	\$1,184	\$1,703	\$519
Canada	\$40,885	\$43,247	\$2,362
Papua New Guinea	\$1,972	\$2,539	\$567
Saudi Arabia	\$50,792	\$53,644	\$2,852
Honduras	\$3,634	\$4,593	\$959

Example: World Bank Poverty Reduction Program

- Future GDP Per Capita

Unit	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$	D_i
Denmark	$41,932	$43,445	$1,513	
Liberia	$806	$878	$72	
Australia	$42,063	$43,544	$1,481	
Afghanistan	$988	$1,946	$958	
South Korea	$31,079	$33,140	$2,061	
Haiti	$1,184	$1,703	$519	
Canada	$40,885	$43,247	$2,362	
Papua New Guinea	$1,972	$2,539	$567	
Saudi Arabia	$50,792	$53,644	$2,852	
Honduras	$3,634	$4,593	$959	

Example: World Bank Poverty Reduction Program

- Future GDP Per Capita

Unit	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$	D_i
Denmark	\$41,932	\$43,445	\$1,513	0
Liberia	\$806	\$878	\$72	1
Australia	\$42,063	\$43,544	\$1,481	0
Afghanistan	\$988	\$1,946	\$958	1
South Korea	\$31,079	\$33,140	\$2,061	0
Haiti	\$1,184	\$1,703	\$519	1
Canada	\$40,885	\$43,247	\$2,362	0
Papua New Guinea	\$1,972	\$2,539	\$567	1
Saudi Arabia	\$50,792	\$53,644	\$2,852	0
Honduras	\$3,634	\$4,593	\$959	1

Example: World Bank Poverty Reduction Program

- Future GDP Per Capita

Unit	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$	D_i
Denmark	\$41,932		\$1,513	0
Liberia		\$878	\$72	1
Australia	\$42,063		\$1,481	0
Afghanistan		\$1,946	\$958	1
South Korea	\$31,079		\$2,061	0
Haiti		\$1,703	\$519	1
Canada	\$40,885		\$2,362	0
Papua New Guinea		\$2,539	\$567	1
Saudi Arabia	\$50,792		\$2,852	0
Honduras		\$4,593	\$959	1

The Fundamental Problem of Causal Inference

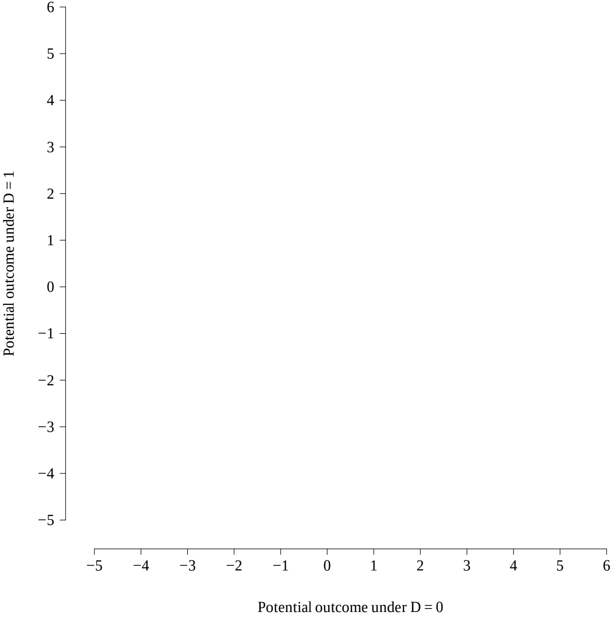
The Fundamental Problem of Causal Inference

- Impossible to observe the values of $Y_i(1)$ and $Y_i(0)$ on the same unit i
- Therefore, impossible to observe the effect of D on unit i
- *A priori* each potential outcome could be observed
- After assignment of D_i, the treatment, one outcome is observed, the other is *counterfactual*

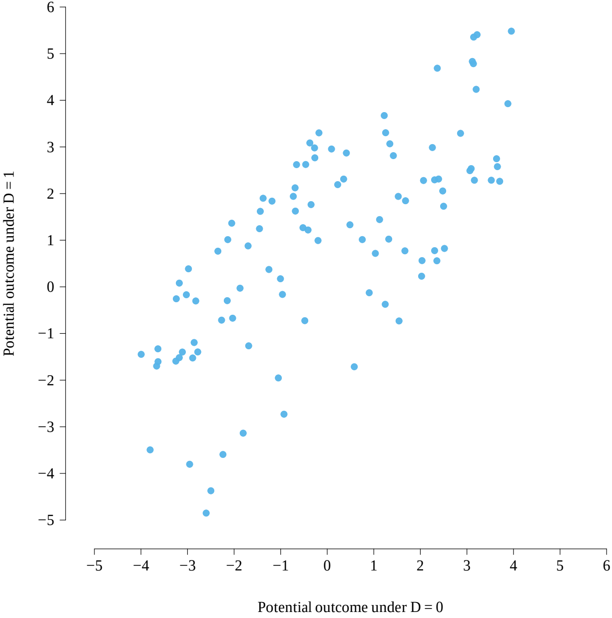
So far in a nutshell

- Causal inference rests on comparing counterfactuals.
- Potential outcomes represent these counterfactuals formally.
- Many types of effects: different comparisons.

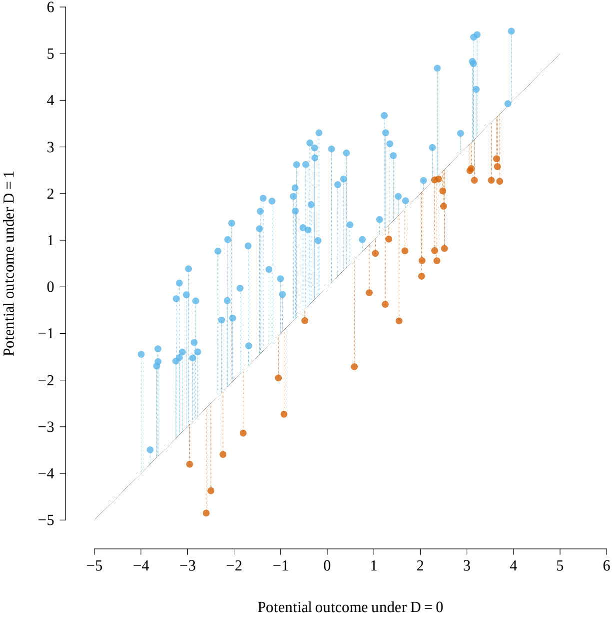
Fundamental Problem of Causal Inference



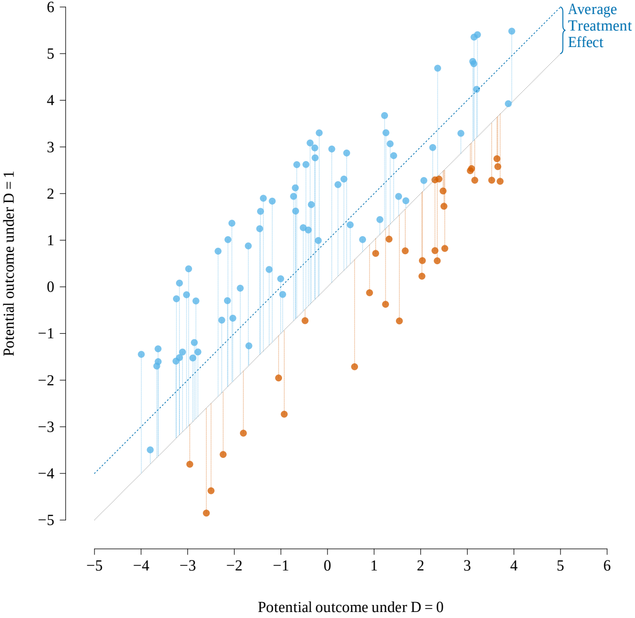
Fundamental Problem of Causal Inference



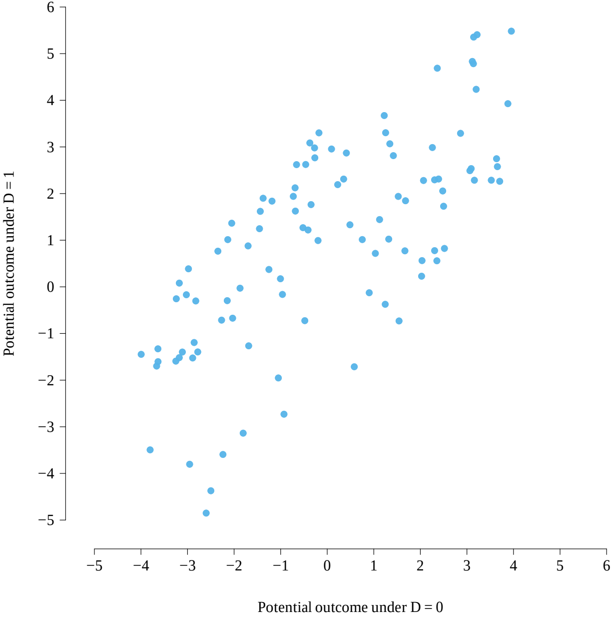
Fundamental Problem of Causal Inference



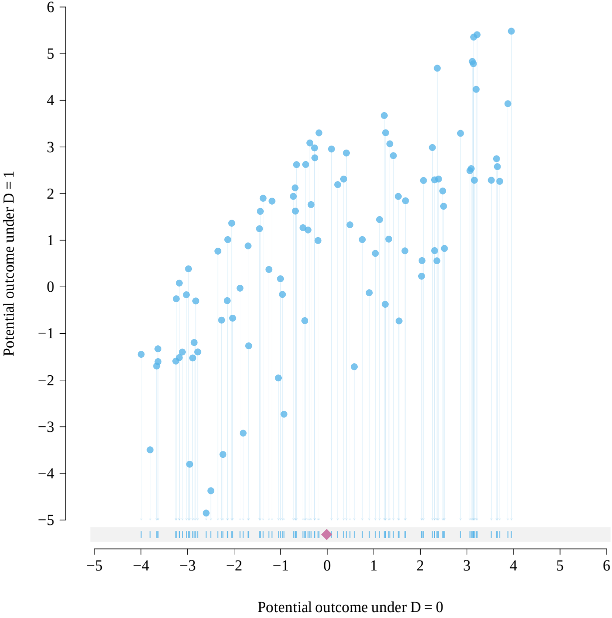
Fundamental Problem of Causal Inference



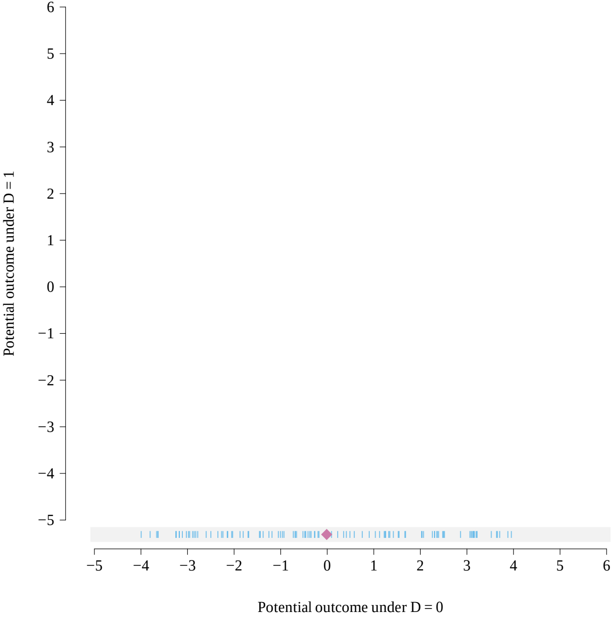
Fundamental Problem of Causal Inference



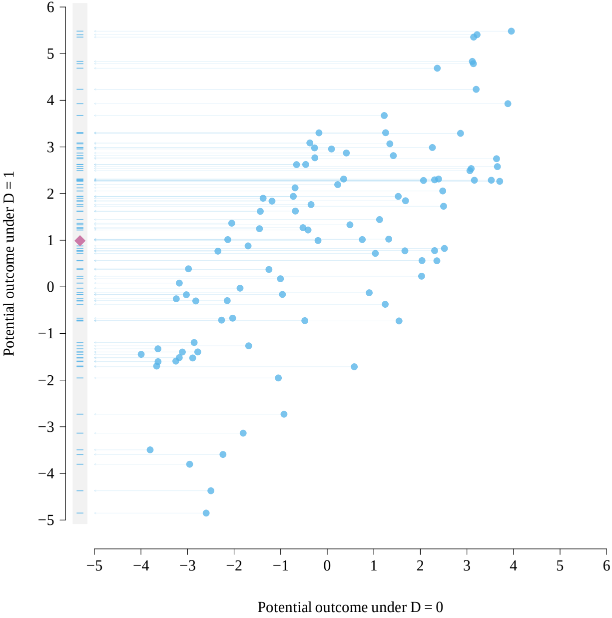
Fundamental Problem of Causal Inference



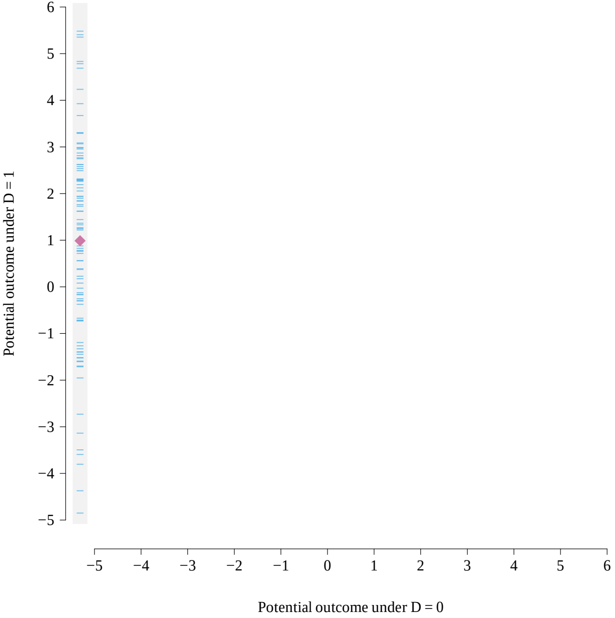
Fundamental Problem of Causal Inference



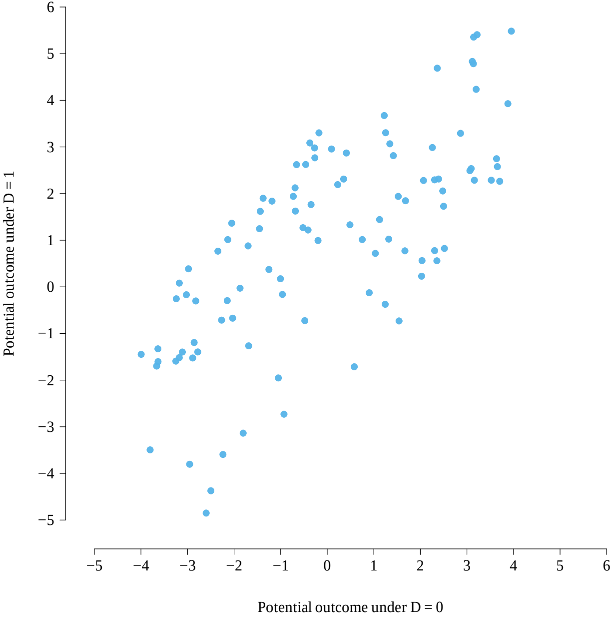
Fundamental Problem of Causal Inference



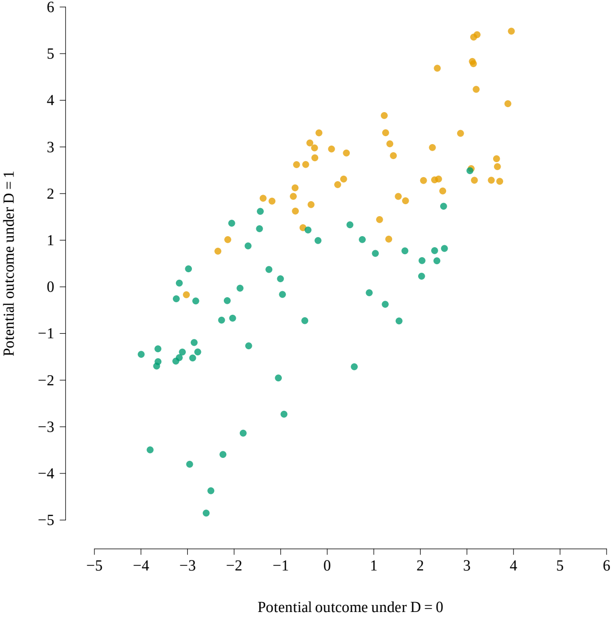
Fundamental Problem of Causal Inference



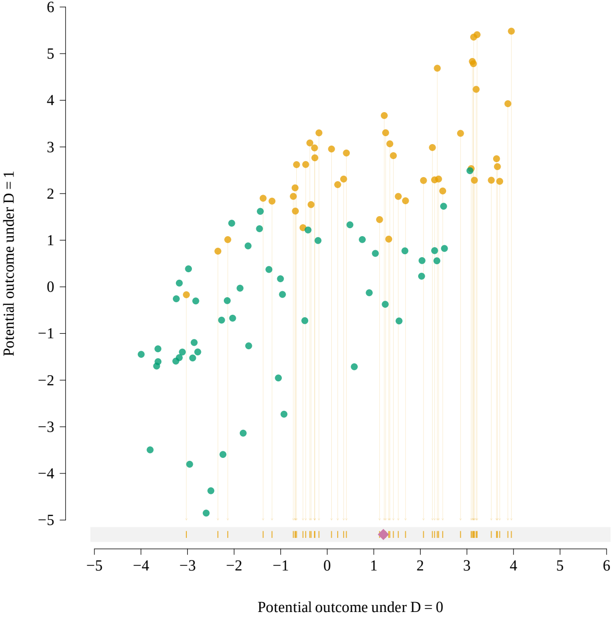
Selection intro treatments



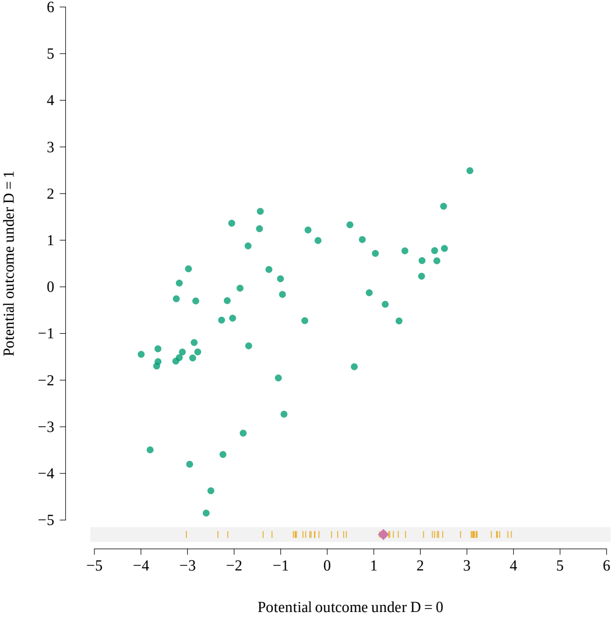
Selection intro treatments



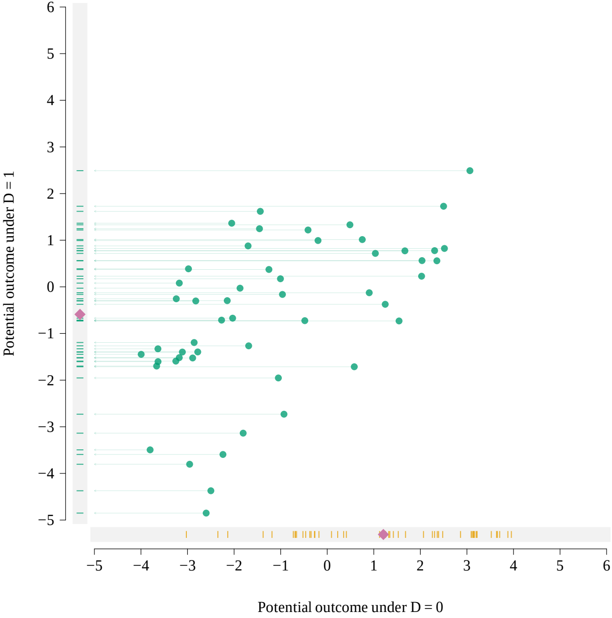
Selection intro treatments



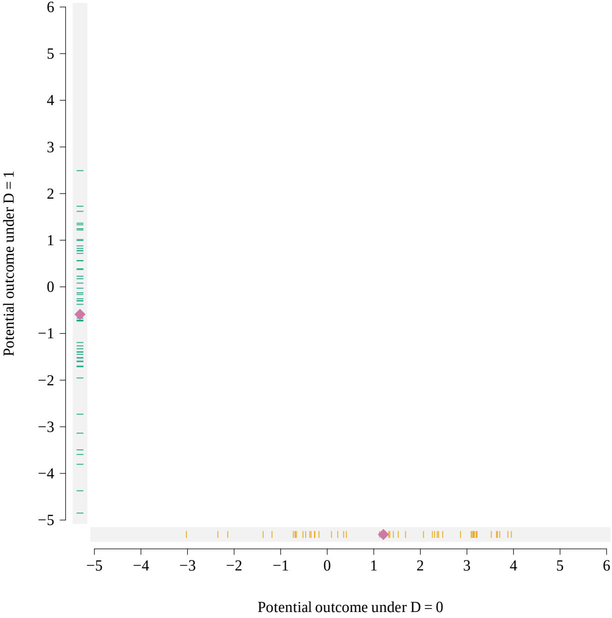
Selection intro treatments



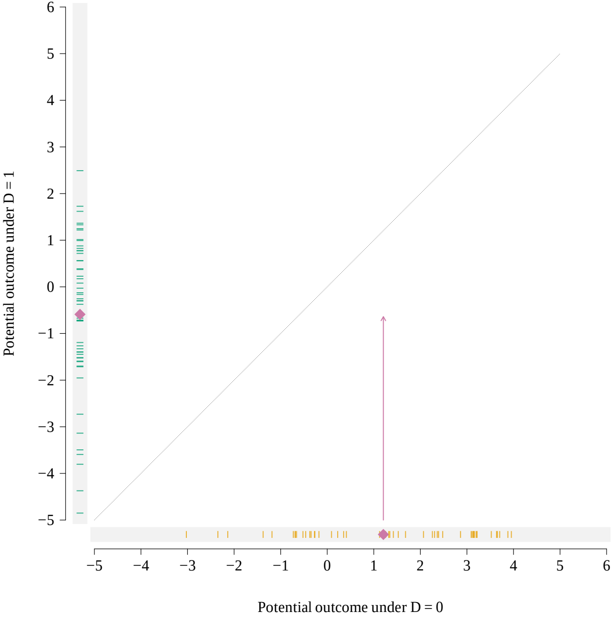
Selection intro treatments



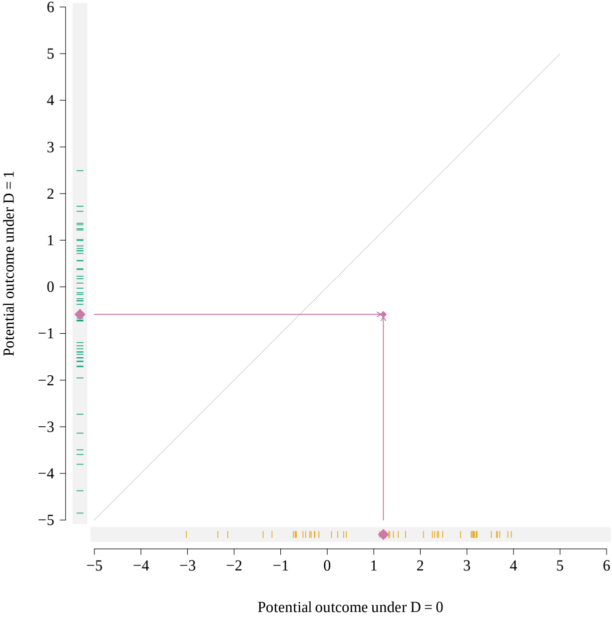
Selection intro treatments



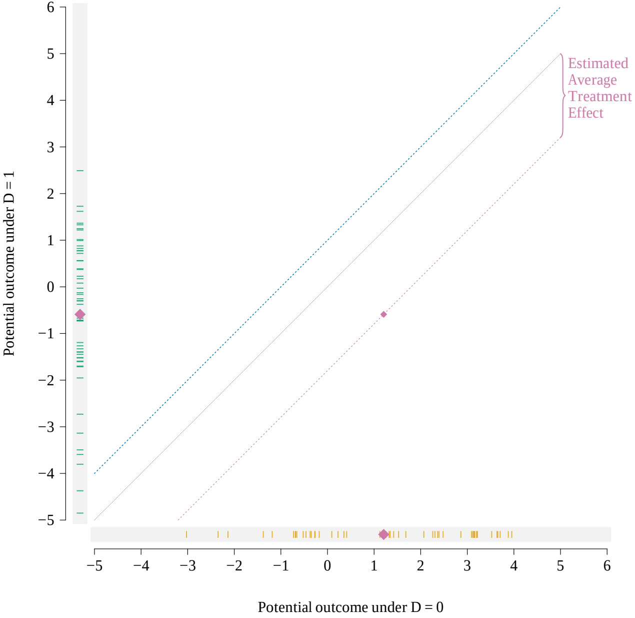
Selection intro treatments



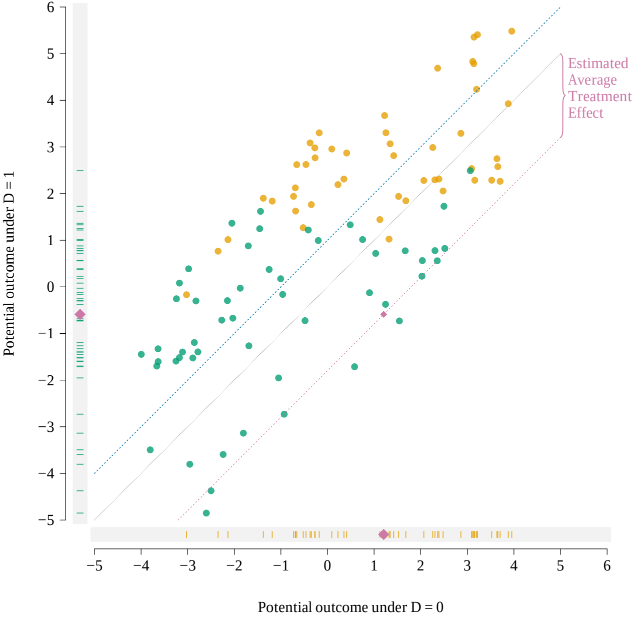
Selection intro treatments



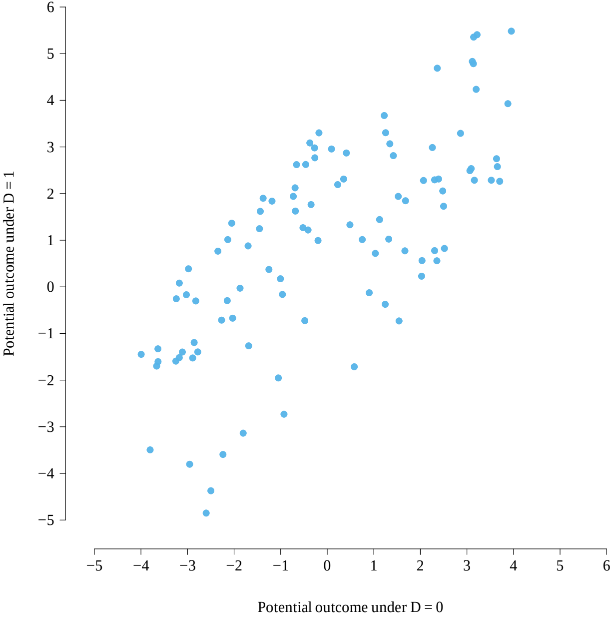
Selection intro treatments



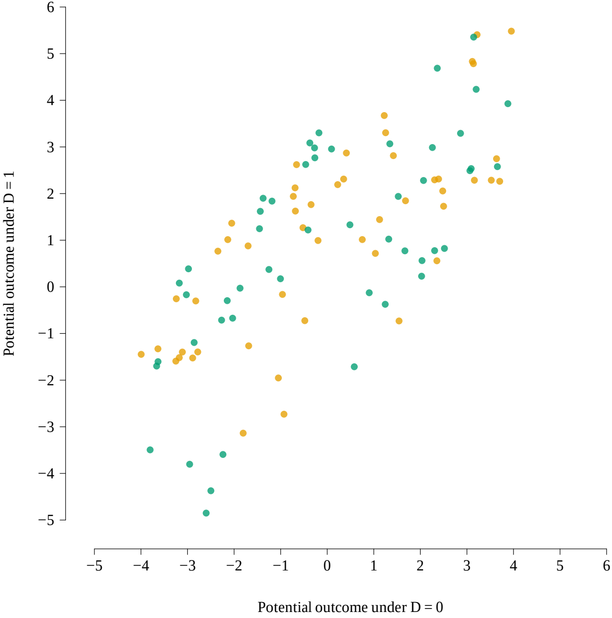
Selection intro treatments



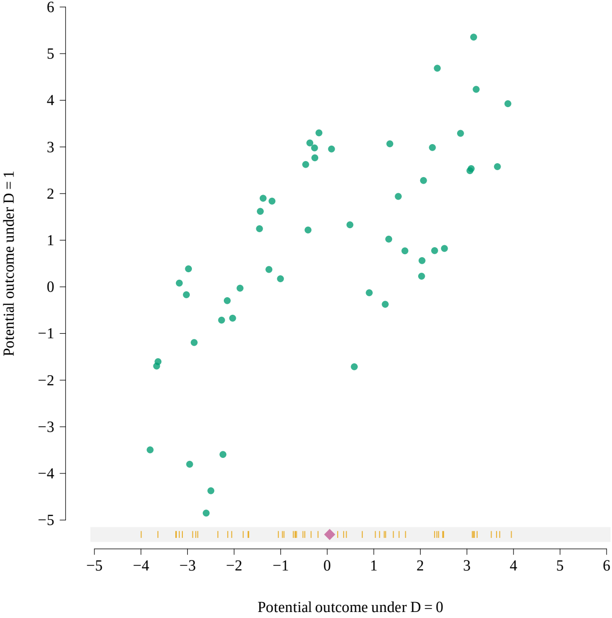
Randomized treatment assignment



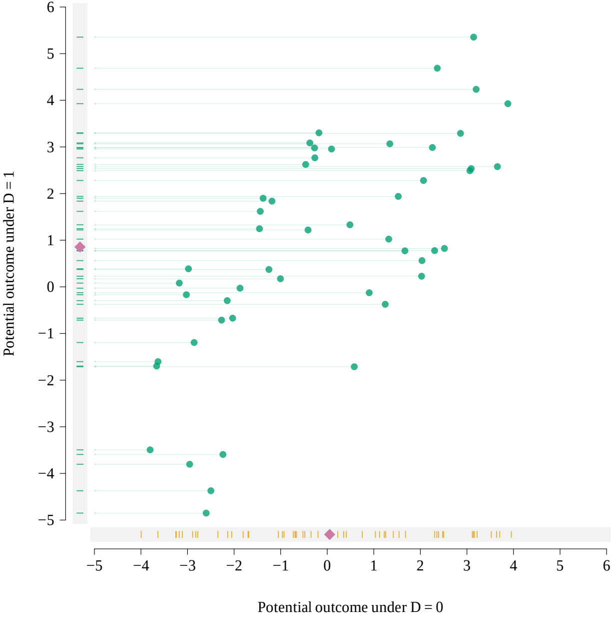
Randomized treatment assignment



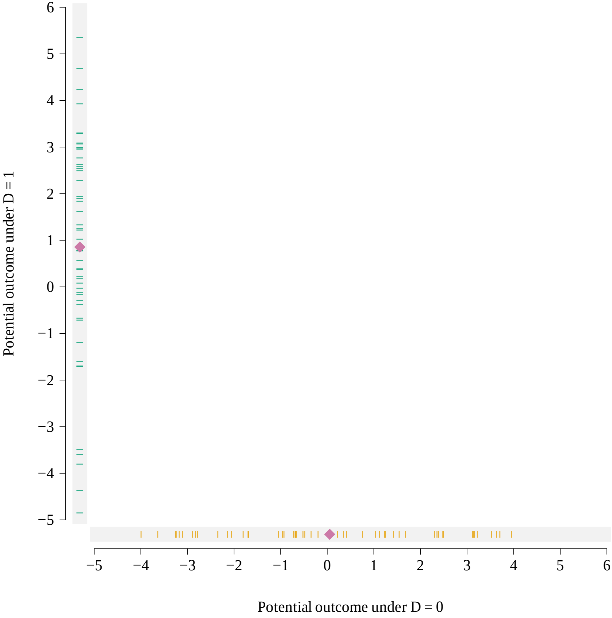
Randomized treatment assignment



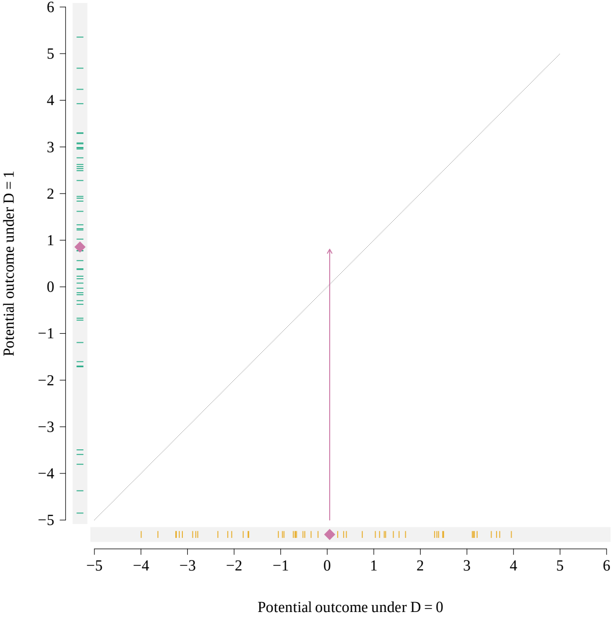
Randomized treatment assignment



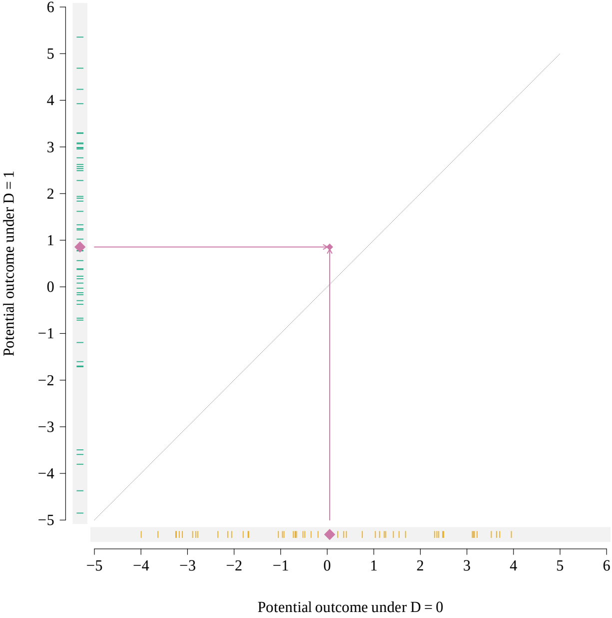
Randomized treatment assignment



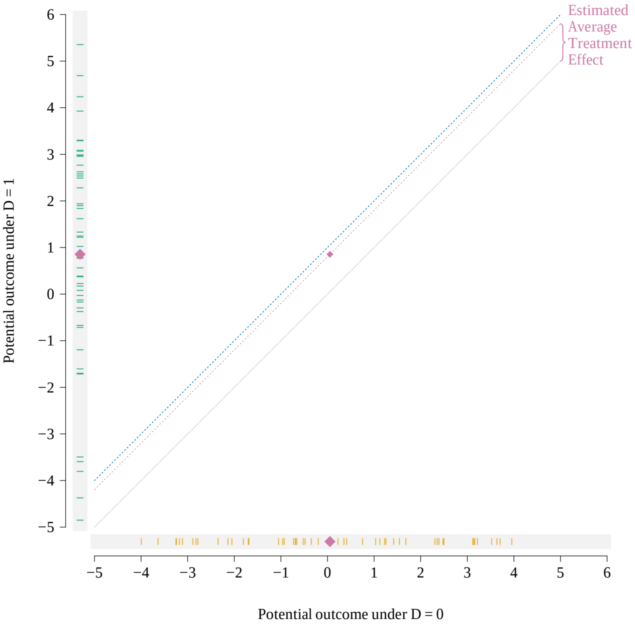
Randomized treatment assignment



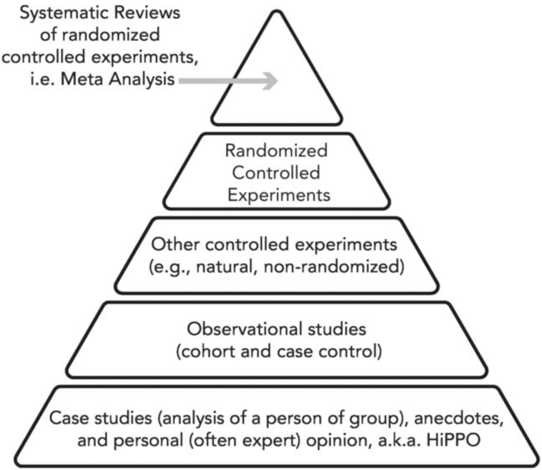
Randomized treatment assignment



Randomized treatment assignment



Randomized experiments as the "gold standard" of causal inference



Source: Kohavi, Tang, and Xu (2020)

The reason that randomized experiments are **the gold standard** for causal inference...

Random assignment

The reason that randomized experiments are **the gold standard** for causal inference...

is that random assignment achieves independence

The reason that randomized experiments are **the gold standard** for causal inference...

is that random assignment achieves independence

(“in expectation” or “asymptotically”).

Random assignment

If we do not have independence, we have *selection bias*

Random assignment

If we do not have independence, we have *selection bias*

- This term means that there is some non-random process that *selects* units into treatment and control groups

Random assignment

If we do not have independence, we have *selection bias*

- This term means that there is some non-random process that *selects* units into treatment and control groups
- If we cannot account for this process, we do not know how to appropriately estimate the relationship between a predictor variable and an outcome

Random assignment

Let's say we don't know the data generating process that led to the observed distribution of D_i

Random assignment

Let's say we don't know the data generating process that led to the observed distribution of D_i

Naïve comparison of average poverty change

$$\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

Random assignment

Let's say we don't know the data generating process that led to the observed distribution of D_i

Naïve comparison of average poverty change

$$\underbrace{\mathbb{E}[Y_i|D_i = 1) - \mathbb{E}[Y_i|D_i = 0)}_{\text{Observed difference in average outcome measures}} \quad = \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

Random assignment

Let's say we don't know the data generating process that led to the observed distribution of D_i

Naïve comparison of average poverty change

$$\underbrace{\mathbb{E}[Y_i|D_i = 1) - \mathbb{E}[Y_i|D_i = 0)}_{\text{Observed difference in average outcome measures}} \quad = \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

$$= \mathbb{E}[Y_i(1)|D_i = 1) - \mathbb{E}[Y_i(0)|D_i = 1] + \mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

Random assignment

Let's say we don't know the data generating process that led to the observed distribution of D_i

Naïve comparison of average poverty change

$$\underbrace{\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]}_{\text{Observed difference in average outcome measures}} = \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

$$= \underbrace{\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 1]}_{\text{Average treatment effect on the treated}} + \mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

Random assignment

Let's say we don't know the data generating process that led to the observed distribution of D_i

Naïve comparison of average poverty change

$$\underbrace{\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]}_{\text{Observed difference in average outcome measures}} = \mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]$$

$$= \underbrace{\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 1]}_{\text{Average treatment effect on the treated}} + \underbrace{\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]}_{\text{Selection bias}}$$

Randomized experiments

- Clever way of avoiding the need to be smart
- Founders: R.A. Fisher and J. Neumann

  ... but randomization as a tool *much* older
- Also known as
  - Randomized Controlled Trials
  - AB Tests

Types of randomized experiments

Lab experiments

Survey experiments

Field experiments

Lab/survey experiments

In a controlled environment ("lab")

- Recruit a sample from the population (ideally representative, but ...)
- Randomly apply a manipulation/treatment
- Hold everything else constant

Can be done

- In a lab
- Within a survey instrument
- Online platforms

Examples

- Do people who receive more praise perform better on tests?
- Do voters who receive more information about candidates tend to like those candidates more?
- Do subjects evaluate female candidates differently from male candidates?

Field experiments

Manipulate variables of interest in natural environment

- Identify a (random) sample in the population
- Randomly assign a manipulation/treatment in the "real world"
- Assume all else is distributed independently of the treatment

Examples

- Do people who receive encouragement to vote actually vote more often?
- Are people who meet a candidate in person more likely to vote?
- Does public shaming cause people to recycle more?

Considerations

- Logistical
- Ethical

Potential problems

Partial Compliance:

- Implementation staff depart from the allocation or treatment procedures
- Subjects end up in a different group than assigned
- Subjects exhibit opposite of compliance (so-called defiers)
- Subjects do not get complete treatment

Attrition

- Subjects drop out or still participate, but cannot be measured/refuse to answer

Potential problems

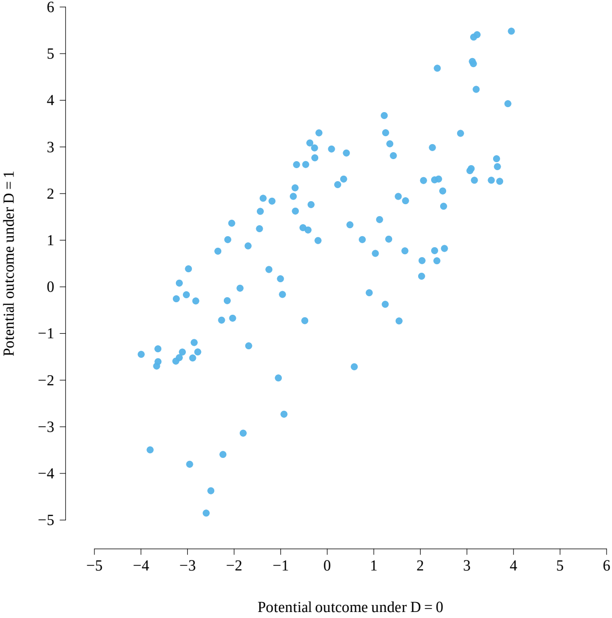
Spillover effects

- Variables unrelated to the experiment have an effect on subjects' behavior
- Treated subjects somehow communicate with untreated subjects

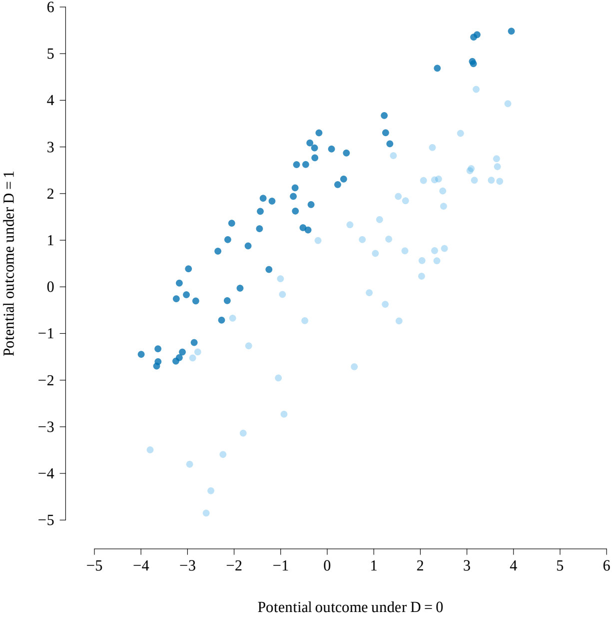
Evaluation-driven effects:

- Hawthorn effects: observation
- John Henry effects: control group works harder
- Demand effects: subjects surmise purpose of the study and act differently than they normally would

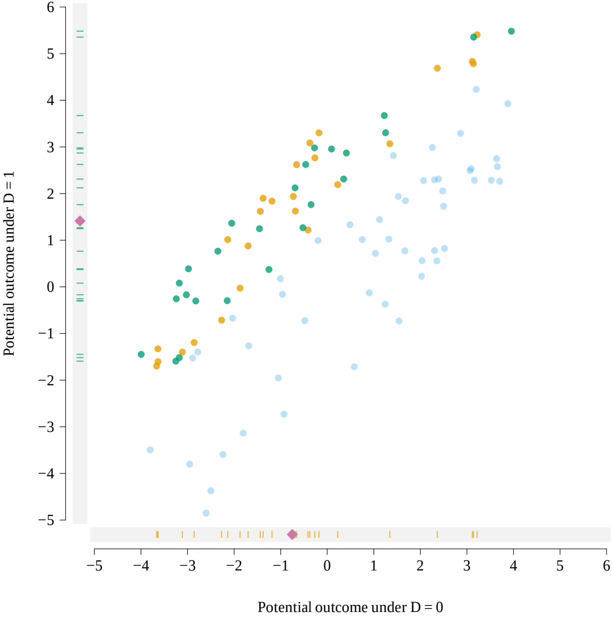
Weird convenience samples



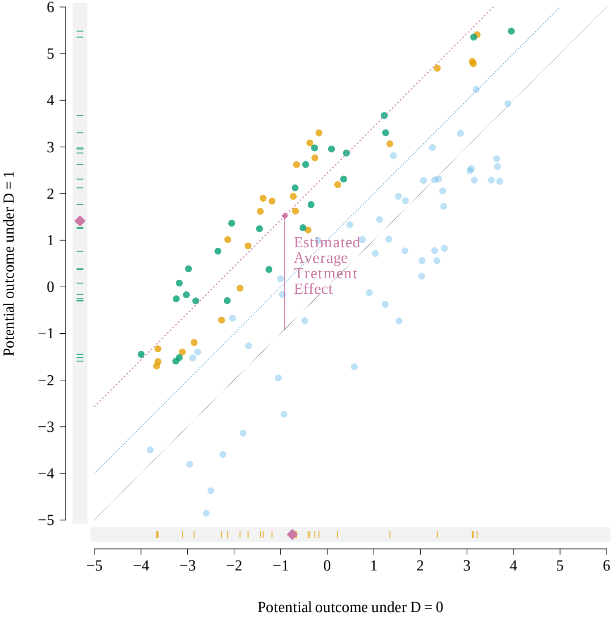
Weird convenience samples



Weird convenience samples



Weird convenience samples



Published: 30 June 2010

Most people are not WEIRD

Joseph Henrich, Steven J. Heine & Ara Norenzayan

Nature **466**, 29 (2010) | Cite this article

To understand human psychology, behavioural scientists must stop doing most of their experiments on Westerners, argue Joseph Henrich, Steven J. Heine and Ara Norenzayan.

Much research on human behaviour and psychology assumes that everyone shares most fundamental cognitive and affective processes, and that findings from one population apply across the board. A growing body of evidence suggests that this is not the case.

Experimental findings from several disciplines indicate considerable variation among human populations in diverse domains, such as visual perception, analytic reasoning, fairness, cooperation, memory and the heritability of IQ[1,2]. This is in line with what anthropologists have long suggested: that people from Western, educated, industrialized, rich and democratic (WEIRD) societies – and particularly American undergraduates – are some of the most psychologically unusual people on Earth[1].

Potential problems

Some social science questions are not possible to answer experimentally

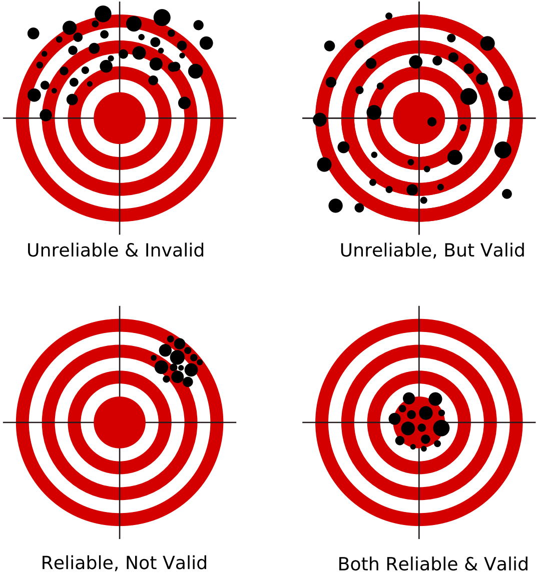
- Cannot randomly assign many variables that we care about

  … e.g. political party, age, country of residence, gender, race

Low external validity

- An experiment is often a poor approximation of the "real world"
- Can we be sure that the relationships we see in an experimental setting carry over to real world politics?
- Especially true with convenience samples

Reliability and validity in randomized experiments



https://en.wikipedia.org/wiki/File:Reliability_and_validity.svg

Reliability and validity in randomized experiments

Validity

- High *internal validity*

Reliability

Validity

- High *internal validity*

  ... but not always doing that well in terms of *external validity*

Reliability

Single randomized experiment, however great, cannot *prove* something.

Validity

- High *internal validity*

  ... but not always doing that well in terms of *external validity*

Reliability

- Randomization introduces noise, i.e. decreases reliability

Single randomized experiment, however great, cannot *prove* something.

Reliability and validity in randomized experiments

Validity

- High *internal validity*

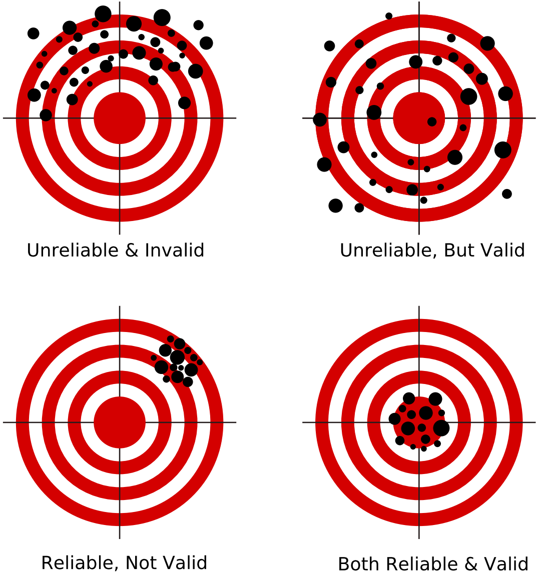
  ... but not always doing that well in terms of *external validity*

Reliability

- Randomization introduces noise, i.e. decreases reliability
- Larger sample sizes increases reliability

Single randomized experiment, however great, cannot *prove* something.

Reliability and validity in randomized experiments



https://en.wikipedia.org/wiki/File:Reliability_and_validity.svg

- Suppose

$$D_i \longrightarrow Y_i$$
$$Y_i(0), Y_i(1) \not\perp\!\!\!\perp D_i$$

that is, there is selection into treatment, and also

$$\boldsymbol{X}_i \longrightarrow D_i$$
$$\boldsymbol{X}_i \longrightarrow Y_i$$

- Intuitively: If we are able to account for all of the $\boldsymbol{X}$ variables that affect both selection into treatment, D, **and** variation in the outcome, Y, then we can claim independence between D and potential outcomes.

- Suppose

$$D_i \longrightarrow Y_i$$
$$Y_i(0), Y_i(1) \not\perp\!\!\!\perp D_i$$

that is, there is selection into treatment, and also

$$\boldsymbol{X}_i \longrightarrow D_i$$
$$\boldsymbol{X}_i \longrightarrow Y_i$$

- Intuitively: If we are able to account for all of the $\boldsymbol{X}$ variables that affect both selection into treatment, D, **and** variation in the outcome, Y, then we can claim independence between D and potential outcomes.
- The big problem: We have to be able to observe and measure all of the X variables that affect selection into treatment and variation in outcomes.

- Suppose

$$D_i \longrightarrow Y_i$$
$$Y_i(0), Y_i(1) \not\perp\!\!\!\perp D_i$$

that is, there is selection into treatment, and also

$$\boldsymbol{X}_i \longrightarrow D_i$$
$$\boldsymbol{X}_i \longrightarrow Y_i$$

- Intuitively: If we are able to account for all of the $\boldsymbol{X}$ variables that affect both selection into treatment, D, **and** variation in the outcome, Y, then we can claim independence between D and potential outcomes.
- The big problem: We have to be able to observe and measure all of the X variables that affect selection into treatment and variation in outcomes.

Next meeting:

Seminar #1

Tue 16 Jan

Bonus

Logistical considerations

- Resource-intensive
- Need subjects ... and therefore, incentives
- Need a place to do it
  - Lab space
  - Computer program
  - Survey distribution
- Usually requires funding
- Only some social science questions lend themselves to experiments

Ethical considerations

What must we do in order to conduct experiments ethically?

- Informed consent
- Beneficence
- Does more good than harm come from the research?
- Respect for anonymity, confidentiality, privacy
- Careful design, competent application, useful results

Deception?

- Is it ethical to mislead participants in order to avoid problems mentioned earlier?

When shouldn't we experiment?

Ethical considerations

INSIDE DEVELOPMENT | TRANSPARENCY AND ACCOUNTABILITY

A new research experiment in Kenya raises questions about ethics

By ***Andrew Green*** *// 08 September 2020*

Water & Sanitation | Research | Infrastructure | World Bank Group | Kenya

A vendor pulls a handcart with empty jerrycans amid a water shortage in Nairobi, Kenya. Photo by: Njeri Mwangi / Reuters

BERLIN — A randomized controlled trial conducted among some of the poorest residents in Nairobi included threatening the disconnection of water and sanitation services if landlords didn't pay outstanding debts.