Policy Evaluation

Lecture 2: What is Causal Inference?

Edmund Malesky, Ph.D.
June 20, 2018
Duke University

Review

- Impact evaluation assesses impact of program on outcome(s)
- Thus, causal inference is central focus of impact evaluation
    - Did program, and program alone, lead to Δ (change) in outcome?
- Correlation ≠ causation warning doesn't satisfy policy makers
    - Seek rationale for decisions: If we do X, will we get Y?
- Causality can be viewed as problem of counterfactual
- Evaluators' main challenge: determine what counterfactual state of world looks like → comparison groups
- Must avoid common pitfalls of invalid comparison groups
- Numerous statistical techniques, but shoe leather also key

Counterfactuals

Causation and Counterfactuals

- Historically, causation defined in terms of observable phenomena
  - e.g., regularity models of Hume (1751) and Mill (1843)
- Now increasingly unified around *counterfactual* as causality
  - Rubin (1974): "Causes are those things that could be treatments in hypothetical experiments."
  - Involves subjunctive conditional statements
    - *"If Maria hadn't received the stipend, she wouldn't be in college"*
- Impact is difference in outcomes *with program* and *without program*, for same individual at same time
- Counterfactual $\longrightarrow$ rule out all outside factors that explain outcome
- But **impossible** to measure same person in two states of world

Causation and Counterfactuals *(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

Causation and Counterfactuals *(Potential Outcomes Framework)*

- Treatment Effect: Y_{1i} - Y_{0i}
  Causal effect of treatment ("treatment effect") on outcome for i is difference between potential outcomes.

- But, we don't see both of these outcomes. The outcomes we actually observe are:

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

  which can be expressed as $Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i)$

- The lack of a counterfactual is the fundamental problem of causal inference. It is a missing data problem.

Causation and Counterfactuals *(The Missing Data Problem)*

i	Y_{1i}	Y_{0i}	Y_i	D_i	$Y_{1i} - Y_{0i}$
1	3	0	3	1	3
2	1	1	1	1	0

Fundamental problem of causal inference

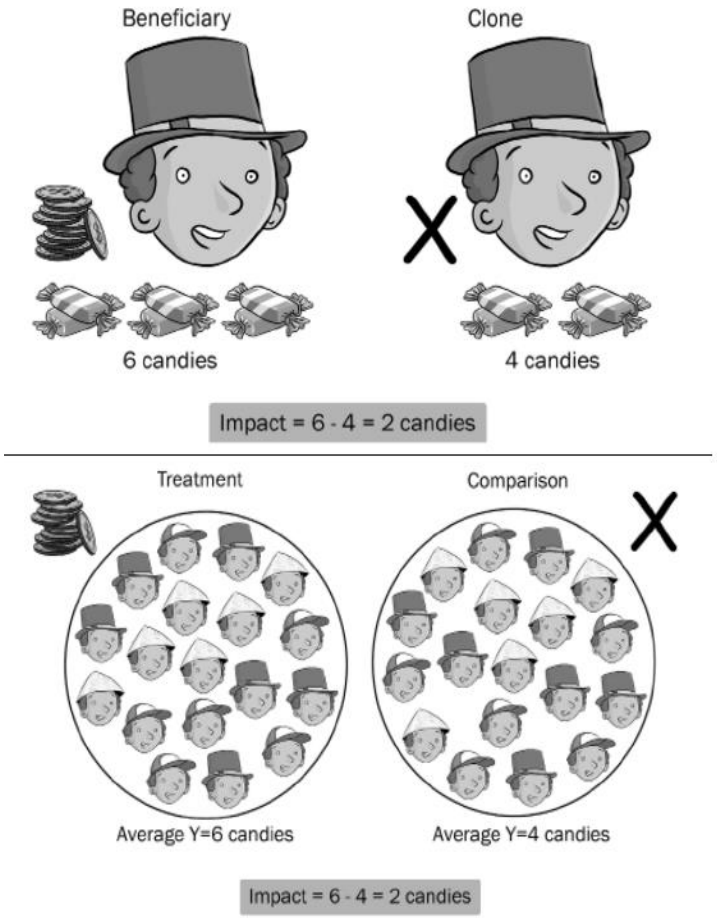
Can never observe both Y_{1i} and Y_{0i}

So, can never know causal effect with certainty

Treatment and Comparison Groups

Evaluator's main challenge: determine what counterfactual state of world looks like

- Easy to estimate outcome under treatment $(Y|D_i=1)$, not $(Y|D_i=0)$ for program participants
- Ideally find perfect clone of program recipient, w/ $D_i=0$
- If you don't have clones, try to create comparison groups.
    - i.e. substitute randomization and sample size for perfect clones.
- Idea that comparison groups, if valid, estimate counterfactual
- **But invalid comparison groups usually lead to biased estimates**



Source: Gertler et al, 2011.

Comparison Groups

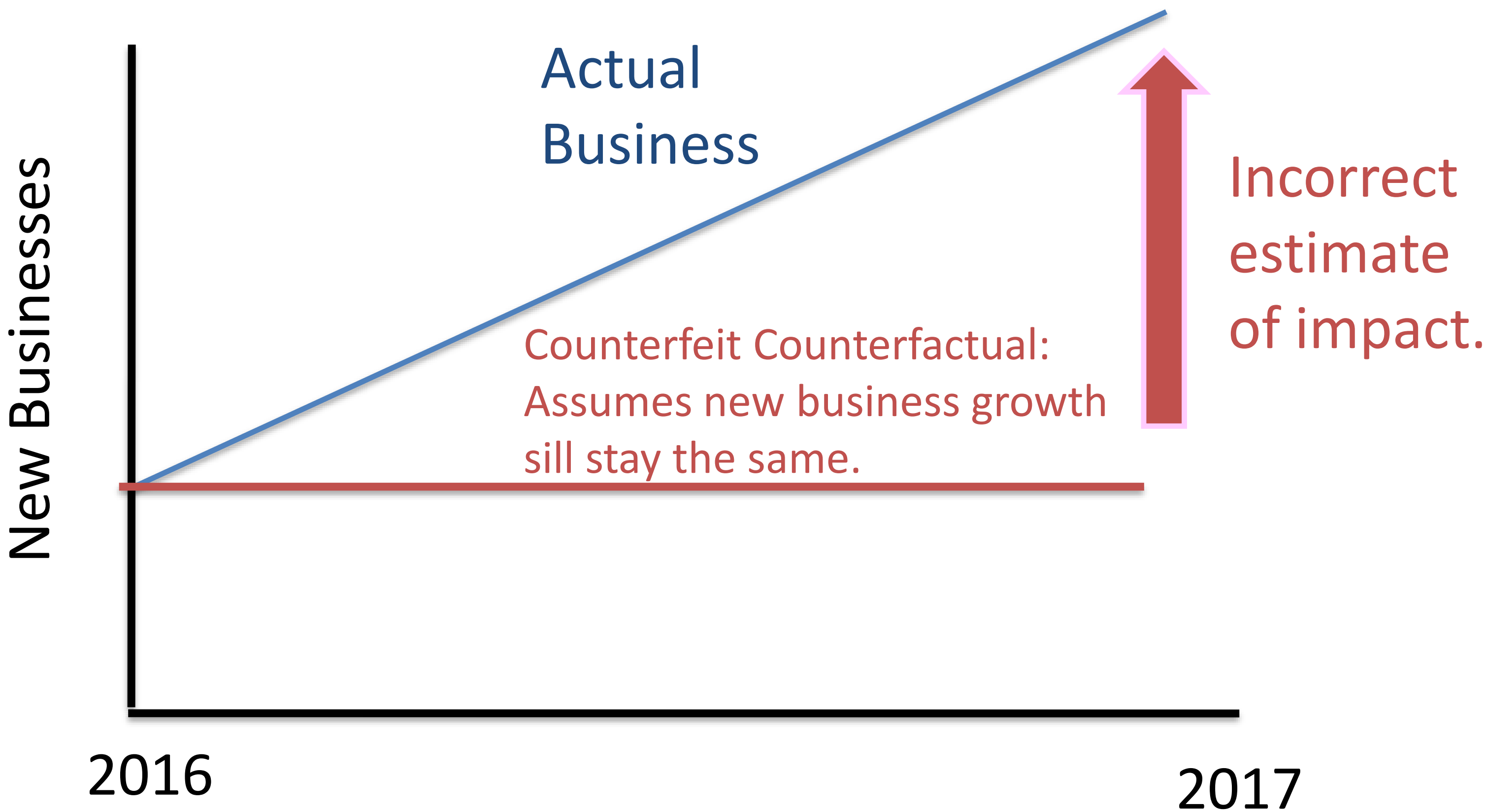
- Key to successful program evaluation: estimate counterfactual by
- finding valid comparison groups
- Treatment & comparison groups need to be same in 3 ways:
  1. **Balanced:** Groups identical in absence of program (on average).
  2. **Parallel Effects:** Groups react to program in same way.
  3. **No Contamination:** Groups not differentially exposed to other interventions during evaluation period.
- Two strategies often used to develop good comparative group
  1. Create comparison group through statistical design
  2. Modify program targeting to erase differences that would have existed between treatment & control

"Counterfeit" Counterfactuals

Beware of invalid comparison groups unlikely to provide unbiased estimates of counterfactuals

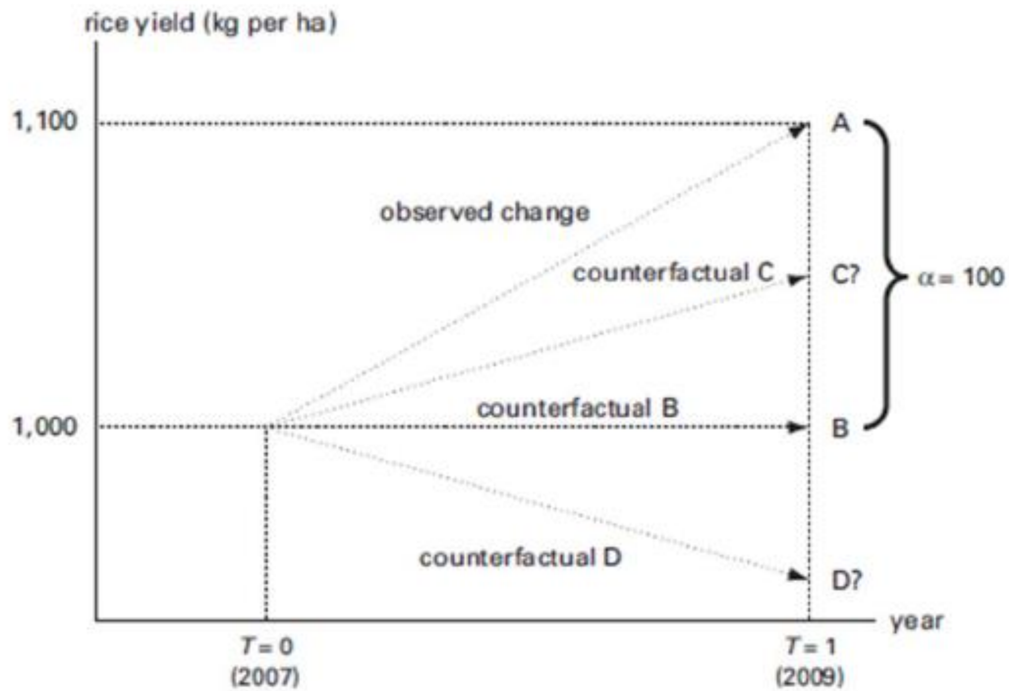
- Two methods particularly likely to give counterfeit counterfactual:
    1. *Comparing outcomes of participants before & after program*
    2. *Compare outcomes of those with & without program*
- If comparison group invalid, then estimates of program effect mixed in with estimates of other differences between groups.

Counterfeit Counterfactual 1: Comparing outcomes of participants *before & after* program



Type 1: Counterfeit Counterfactual

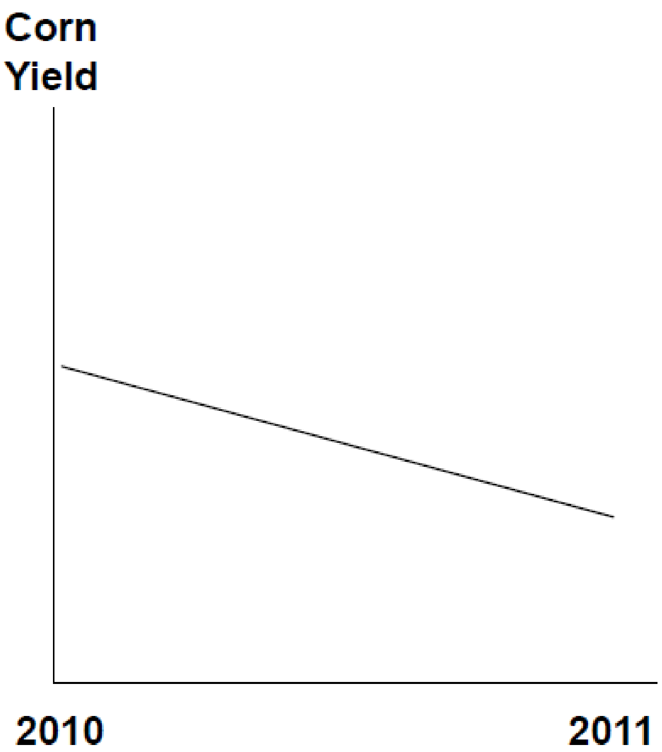
- "Reflexive method" tracks Δ in outcome of participants over time
- Assumes w/out program, outcome would be same as pre-program
- Rarely holds, so poor estimate of counterfactual
- May under or overestimate impact
- Can *control* for some factors that affect outcome, but unobservables remain
- Helps show program objectives met, but can't attribute Δ to project



Type 1: Counterfeit Counterfactual

(Fertilizer Yield Program)

- Fertilizer program targets poor region (A) of country
- To receive fertilizer, farmers enroll at local office
- Starts in 2010; ends in 2011
- We observe *decrease* in yields among recipients during program
- Did program fail?
  - No, there was a national drought.
  - Failure of reflexive comparison



Type 1: Counterfeit Counterfactual

(Health Insurance Program)

- How much did health expenditures fall for poor w/ insurance subsidy?
  If $9, then donors will expand nationally
- Find statistically significant difference before/after using reflexive method
  But not $9. Should program be expanded?
- Now control for various other factors. Should program be expanded?

Table 3.1 Case 1—HISP Impact Using Before-After (Comparison of Means)

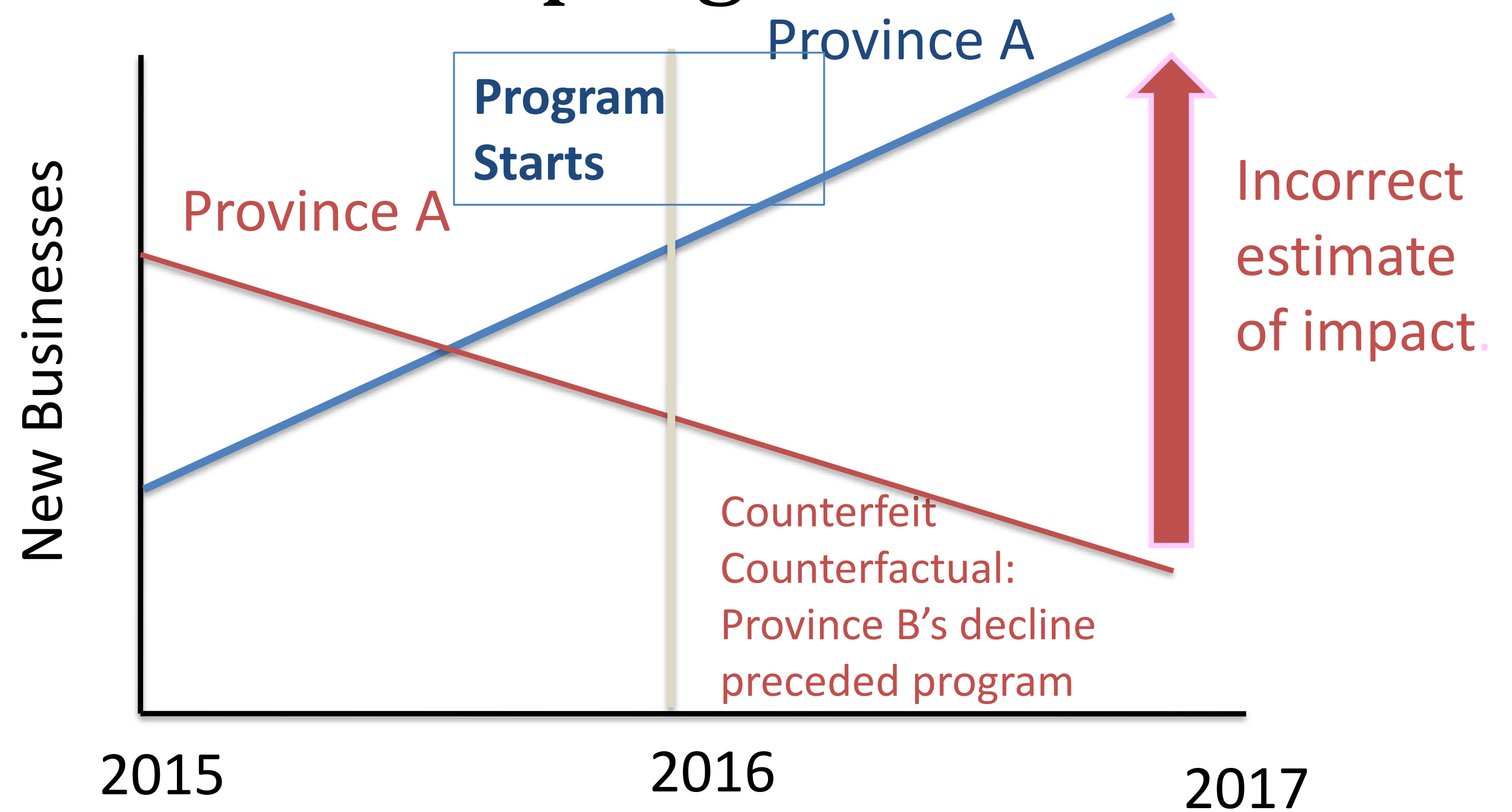
	After	Before	Difference	*t*-stat
Household health expenditures	7.8	14.4	−6.6	−28.9

Table 3.2 Case 1—HISP Impact Using Before-After (Regression Analysis)

	Linear regression	Multivariate linear regression
Estimated impact on household health expenditures	−6.59** (0.22)	−6.65** (0.22)

Source: Gertler et al, 2011.

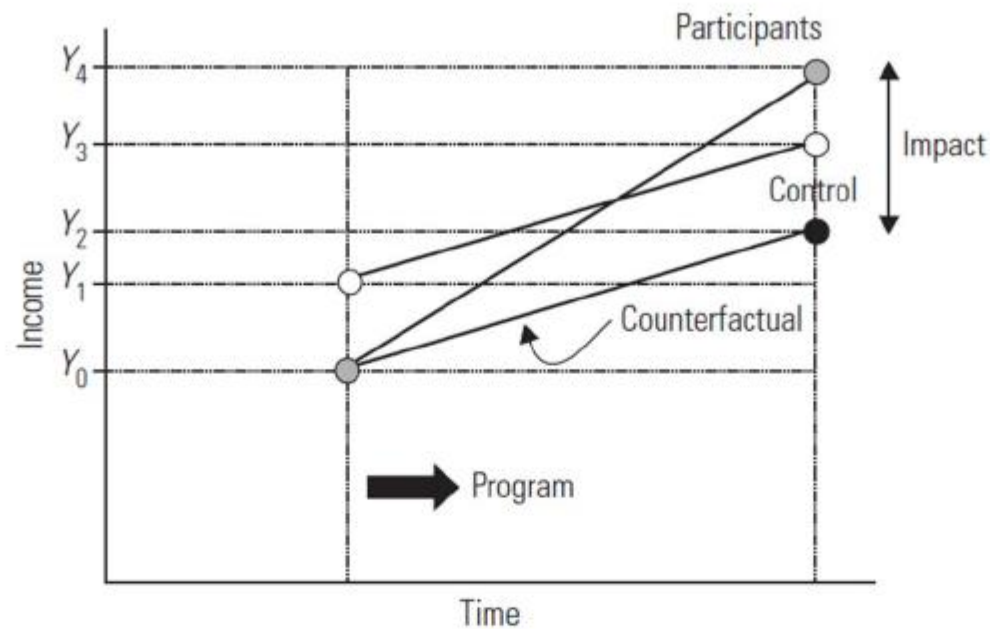
Counterfeit Counterfactual 2: Compare outcomes of those *with* & *without* program



Type 2: Counterfeit Counterfactual

Another method: compare outcomes of those *with & without* program

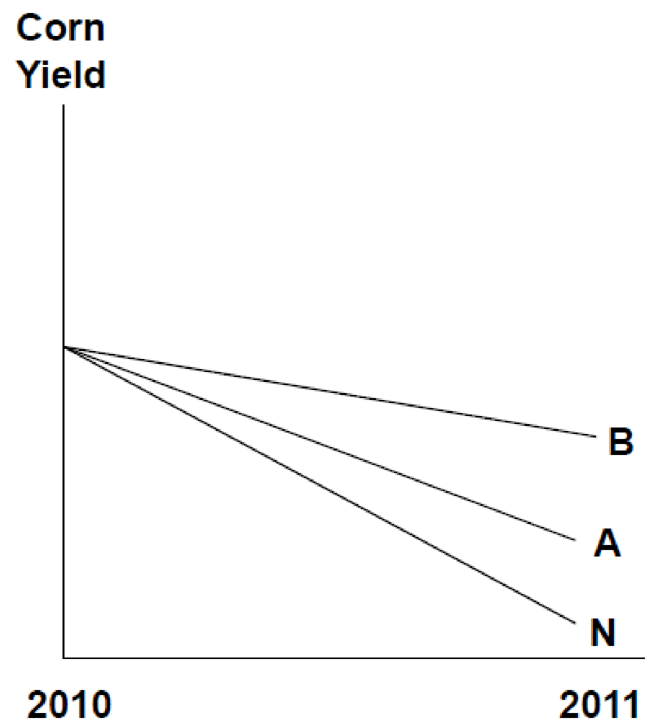
- Often poor estimate of counterfactual due to *selection bias*
- Programs are targeted, so intended differences by design
- Self-selection another key issue, as participation is typically voluntary
- Treatment & comparison groups vary, in both observable & unobservable ways



Type 2: Counterfeit Counterfactual

(Fertilizer Yield Program)

- Compare recipients to farmers in B
- Recipients decline > region B farmers.
  - Negative program impact?
- *Maybe just program placement*
  - *Better soil or irrigation in B*
- What if decline < neighbors?
  - Positive program impact?
- *Maybe farmers w/ greater ability enroll, & can survive drought better*
- What if decline = nonrecipient neighbors?
  - No program impact?
- *Maybe spillovers of fertilizer*



Type 2: Counterfeit Counterfactual
(Health Insurance Program)

- Compare enrolled and non-enrolled families in country
- Find statistically significant difference. *Should program be expanded?*
- Now control for various other factors. *Should program be expanded?*

Table 3.3 Case 2—HISP Impact Using Enrolled-Nonenrolled (Comparison of Means)

	Enrolled	Nonenrolled	Difference	*t*-stat
Household health expenditures	7.8	21.8	−13.9	−39.5

Table 3.4 Case 2—HISP Impact Using Enrolled-Nonenrolled (Regression Analysis)

	Linear regression	Multivariate linear regression
Estimated impact on household health expenditures	−13.9** (0.35)	−9.4** (0.32)

Source: Gertler et al, 2011.

Comparing Treated and Untreated

- Comparing outcomes for the treated and untreated often yields incorrect estimates
  $E[Y|D = 1] - E[Y|D = 0]$
  $\mathrm{E}[\mathrm{Y}_1|D = 1] - E[Y_0|D = 0]$
  $\mathrm{E}[\mathrm{Y}_1|D = 1] - E[Y_0|D = 0] + (E[Y_0|D = 1] - E[Y_0|D = 1])$
  which rearranges to:
  $\mathrm{E}[\mathrm{Y}_1|D = 1] - E[Y_0|D = 1] + E[Y_0|D = 1] - E[Y_0|D = 0]$
  which consists of the Average Treatment Effect on the Treated (ATET) and Selection Bias:
  $$\underbrace{E[Y_1 - Y_0|D = 1]}_{\text{ATET}}$$
  $$+ \underbrace{E[Y_0|D = 1] - E[Y_0|D = 0]}_{\text{Selection Bias}}$$

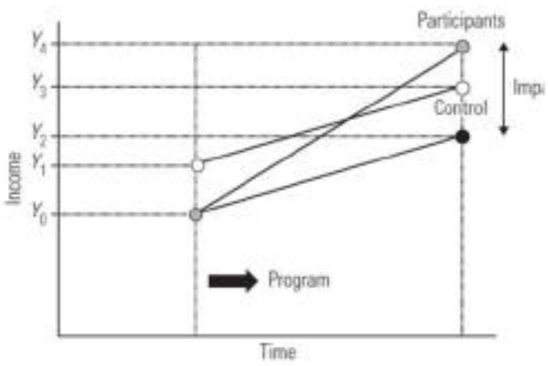
Comparing Treated and Untreated (An Example)

TABLE 1.—MEAN EARNINGS PRIOR, DURING, AND SUBSEQUENT TO TRAINING FOR 1964 MDTA CLASSROOM TRAINEES AND A COMPARISON GROUP

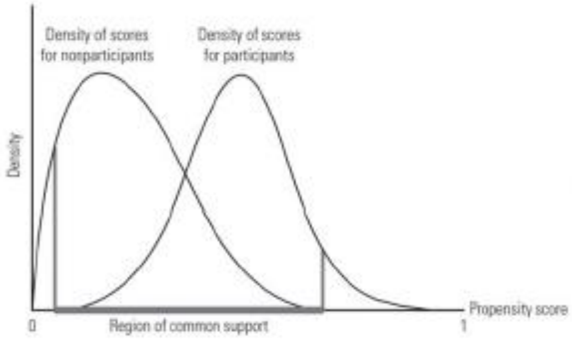
	White Males		Black Males		White Females		Black Females	
	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group	Trainees	Comparison Group
1959	$1,443	$2,588	$ 904	$1,438	$ 635	$ 987	$ 384	$ 616
1960	1,533	2,699	976	1,521	687	1,076	440	693
1961	1,572	2,782	1,017	1,573	719	1,163	471	737
1962	1,843	2,963	1,211	1,742	813	1,308	566	843
1963	1,810	3,108	1,182	1,896	748	1,433	531	937
1964	1,551	3,275	1,273	2,121	838	1,580	688	1,060
1965	2,923	3,458	2,327	2,338	1,747	1,698	1,441	1,198
1966	3,750	4,351	2,983	2,919	2,024	1,990	1,794	1,461
1967	3,964	4,430	3,048	3,097	2,244	2,144	1,977	1,678
1968	4,401	4,955	3,409	3,487	2,398	2,339	2,160	1,920
1969	$4,717	$5,033	$3,714	$3,681	$2,646	$2,444	$2,457	$2,133
Number of Observations	7,326	40,921	2,133	6,472	2,730	28,142	1,356	5,192

Impact Evaluation Approaches

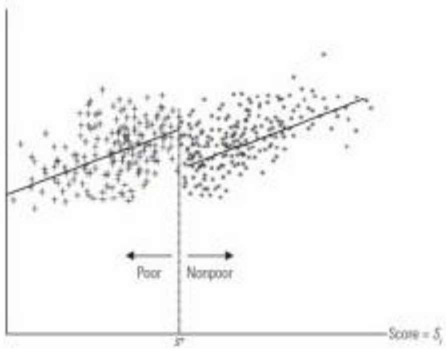
Differences in Differences



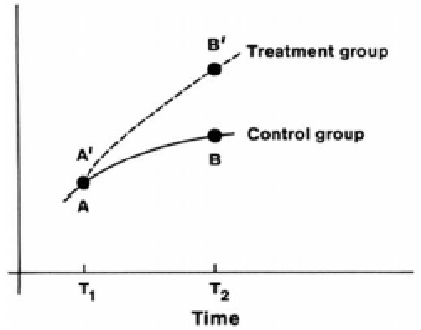
Matching



Regression Discontinuity

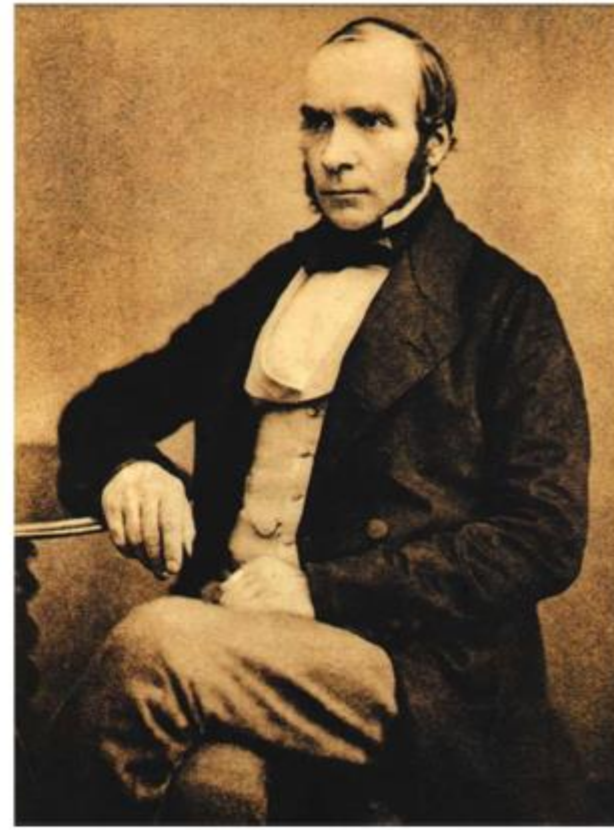


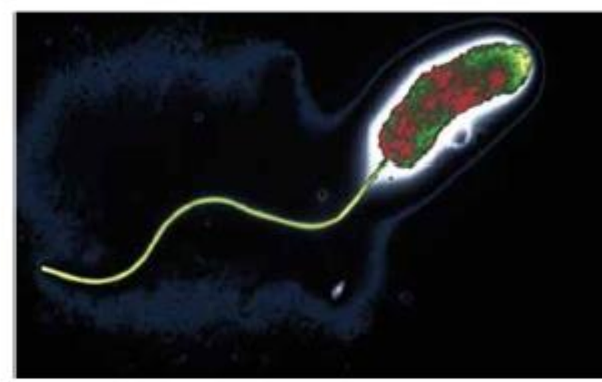
Randomized Evaluation



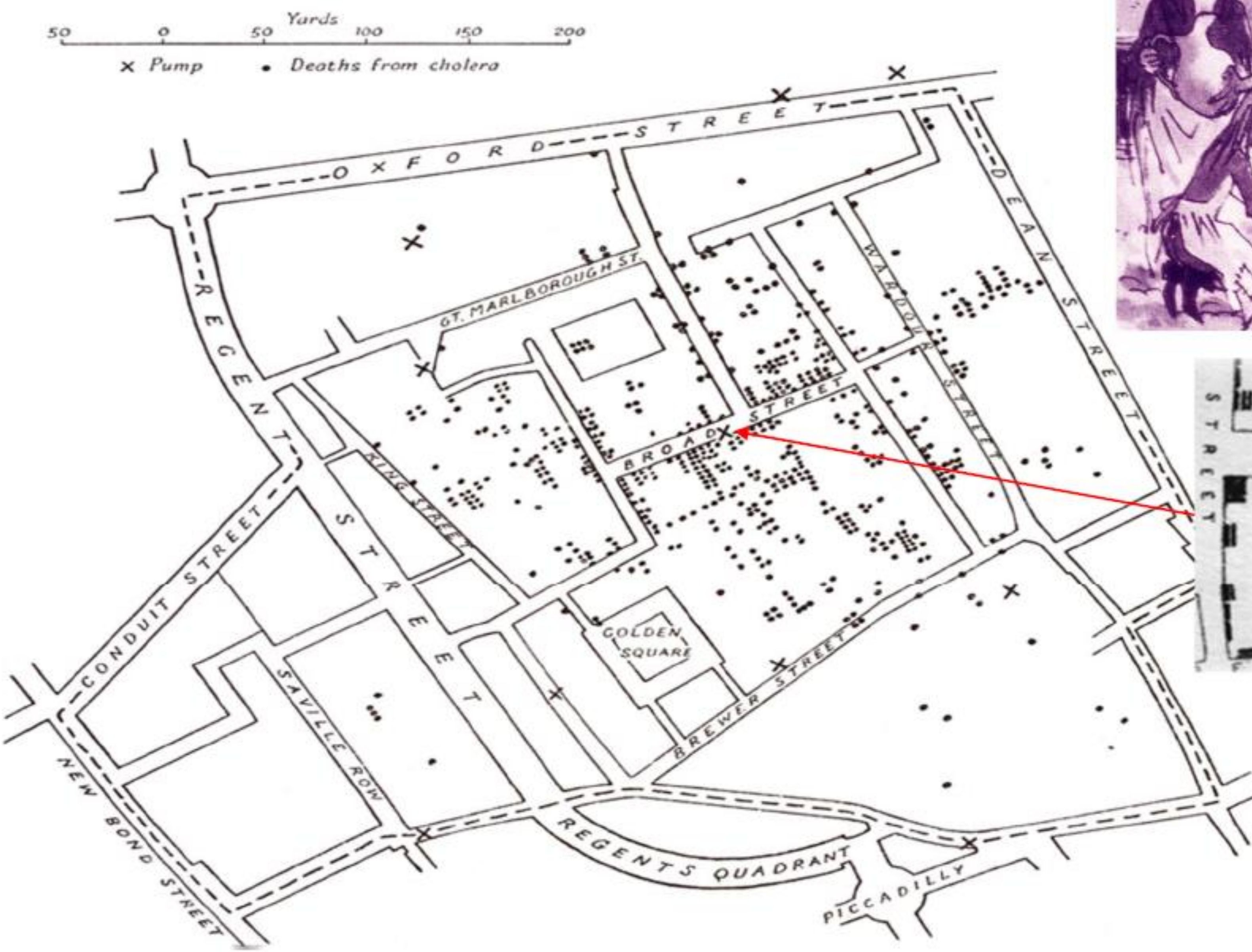
John Snow and the Discovery of Cholera

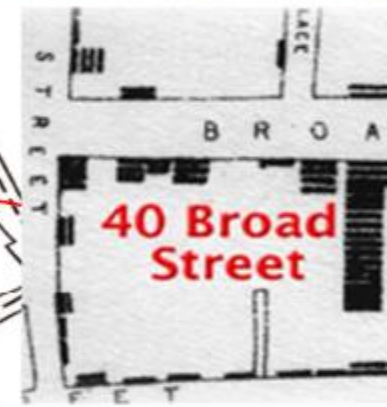
- Dr. John Snow (1813-1858) one of founders of modern epidemiology.
- Studied London Cholera Outbreak in 1848-1854.
- 250,000 cases & 53,000 deaths in peak two-year period.
- Didn't know germs caused disease, airborne "miasmas" common belief.
- Snow hypothesis: causal agent in sewage-contaminated water





The Broad Street Pump



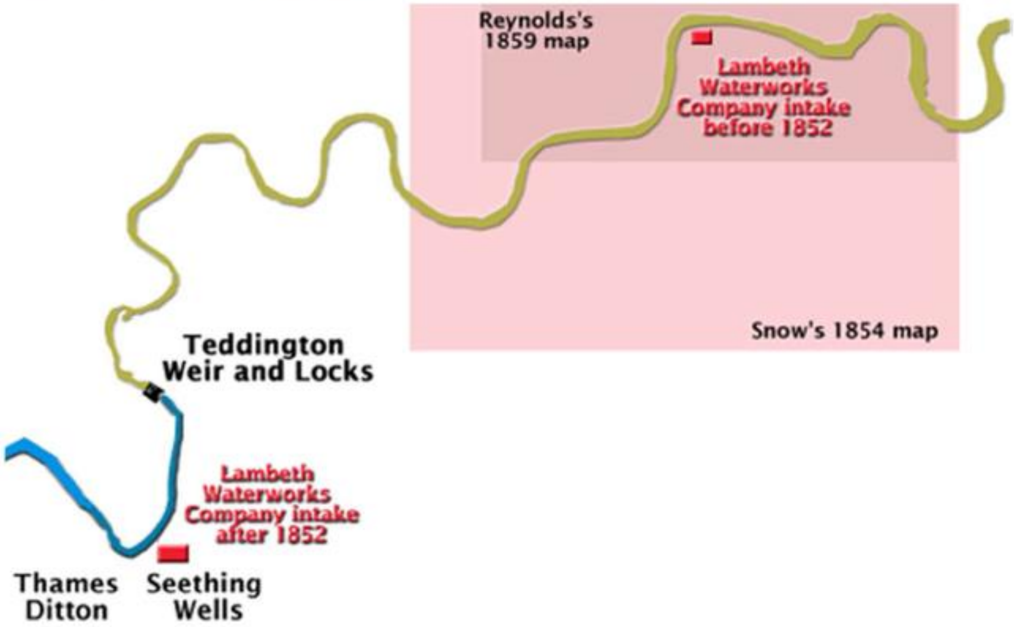


Water Supply in London

Lambeth Water Company

Before 1852: intake from Thames River in London → sewage

In 1852: moved intake 22 miles up river → no sewage

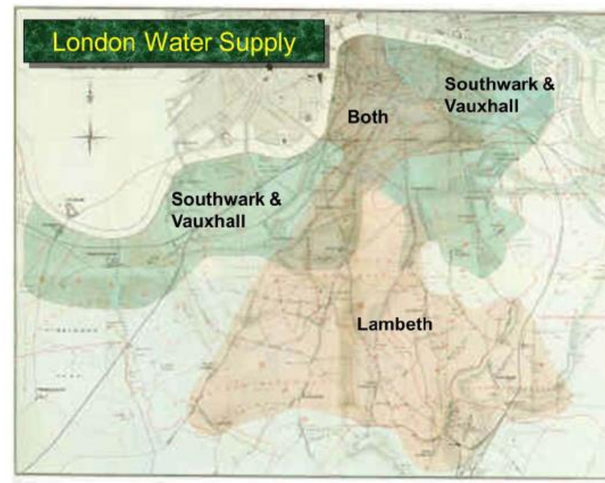


Southwark and Vauxhall Water Company

Continued to draw water from contaminated part of Thames River

- **Natural experiment:** Lambeth company has clean water. Southwak and Vauxhall company have infected sources.
- Location of pumps is “as if” random. Broad Street pump is S&V. Clean pumps are Lambeth. Residents do not get to select company (no selection bias).

The mixing of the (water) supply is of the most intimate kind. The pipes of each Company go down all the streets, and into nearly all the courts and alleys. A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time when the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different Companies…It is obvious that no experiment could have been devised which would more thoroughly test the effect of water supply on the progress of cholera than this” (Snow 1855: 74-75).



- For houses served by Southwark and Vauxhall, the death rate from cholera was 315 per 10,000; for houses served by Lambeth, it was a mere 37 per 10,000

TABLE VI.

Sub-Districts.	Population in 1851.	Deaths from Cholera in 1853.	Deaths by Cholera in each 100,000 living.	Water Supply.
St. Saviour, Southwark	19,709	45	227	Southwark and Vauxhall Water Company only.
St. Olave	8,015	19	237	
St. John, Horsleydown	11,360	7	61	
St. James, Bermondsey	18,899	21	111	
St. Mary Magdalen	13,934	27	193	
Leather Market	15,295	23	153	
Rotherhithe*	17,805	20	112	
Wandsworth	9,611	3	31	
Battersea	10,560	11	104	
Putney	5,280	—	—	
Camberwell	17,742	9	50	
Peckham	19,444	7	36	
Christchurch, Southwk.	16,022	7	43	Lambeth Water Company, and Southwark and Vauxhall Company.
Kent Road	18,126	37	204	
Borough Road	15,862	26	163	
London Road	17,836	9	50	
Trinity, Newington	20,922	11	52	
St. Peter, Walworth	29,861	23	77	
St. Mary, Newington	14,033	5	35	
Waterloo (1st part)	14,088	1	7	
Waterloo (2nd part)	18,348	7	38	
Lambeth Church (1st part)	18,409	9	48	
Lambeth Church (2nd part)	26,784	11	41	
Kennington (1st part)	24,261	12	49	
Kennington (2nd part)	18,848	6	31	
Brixton	14,610	2	13	
Clapham	16,290	10	61	
St. George, Camberwell	15,849	6	37	
Norwood	3,977	—	—	Lambeth Water Company only.
Streatham	9,023	—	—	
Dulwich	1,632	—	—	
First 12 sub-districts	167,654	192	114	Southwk. & Vaux.
Next 16 sub-districts	301,149	182	60	Both Companies.
Last 3 sub-districts	14,632	—	—	Lambeth Comp.

* A part of Rotherhithe was supplied by the Kent Water Company; but there was no cholera in this part.

Evidence

- Striking difference in death rate varied by water source
  - Not selected by current residents, typically unknown
- Snow: >1,000 lives would be saved if Southwark moved intake
- Example of how shoe leather—rather than reliance on statistical technnology—can achieve causal inference w/ observational data

	Number of Houses	Deaths from Cholera	Deaths Per 10,000 Houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

Deviant Cases Prove Rule

- Brewery workers around Broad Street did not get sic (sick).
- At addresses closer to Lambeth pumps, infected people preferred Broad Street water (i.e. a widow sent out for it).

SUMMING UP

10 Things You Need to Know About Causal Inference

1. A causal claim is a statement about what didn't happen.

2. There is a fundamental problem of causal inference.

3. You can estimate average causal effects even if you cannot observe any individual causal effects.

4. If you know that, on average, A causes B and B causes C, this does not mean that you know that A causes C.

10 Things You Need to Know About Causal Inference

5. The counterfactual model is all about contribution, not attribution.

6. X can cause Y even if there is no “causal path” connecting X and Y.

7. Correlation is not causation

8. X can cause Y even if X is not a necessary condition or a sufficient condition for Y.
9. Estimating average causal effects does not require that treatment and control groups are identical.

10. There is no causation without manipulation

10 Things You Need to Know About Causal Inference

9. Estimating average causal effects does not require that treatment and control groups are identical.

10. There is no causation without manipulation