# Lecture 12: Causality and Experiments: II
## Modeling Social Data, Spring 2017
## Columbia University

Wanting Wang

April 28, 2017

- Random Assignment

$$\overline{\underline{T_i \;|C_i}}$$

$$\nearrow \qquad \nwarrow$$

what happens under treatment — what happens under control

Theoretically, in reality, you can never measure them at the same time.

Treatment group: $\overline{\underline{T_i \;||||}}$

Average treatment outcome: $\hat{\bar{\mathrm{T}}} = \frac{1}{N_T}\sum_i T_i$

Control group: $\overline{\underline{|||||C_i}}$

Average control outcome: $\hat{\bar{\mathrm{C}}} = \frac{1}{N_C}\sum_i C_i$

Average of random sample is an unbiased estimate, and the average treatment effect is: A$\hat{T}E = \hat{\bar{\mathrm{T}}} - \hat{\bar{\mathrm{C}}}$

- Problems
  - Small sample size: large sample size can reduce SE and lower the chance of the estimate being way too off.
  - Researcher degrees of freedom: tend to utilize various method to "mine" the data for a nice result. The hypotheses and analysis method should be set before touching the data.
  - Publication bias: the reproducibility is questionable, especially in a field where the power is relatively low.
  - P-hacking

- Power Analysis
  N = sample size
  $\alpha$ = significance level = $P$(significance |no effect)
  power = 1 - $\beta$ = $P$(significance |effect) $\Longleftarrow$ chance of detecting a real effect if one exists
  - How to get hypothesized p-value? Run a pilot study!

- Limitation
  - Sometimes isn't feasible/ethical
  - Costly in terms of time and money
  - Difficult to create convincing parallel world
  - People inevitably deviate from assignment
- Non-Compliance

| | | |
|---|---|---|
| $T_i \ \vert C_i$ | Compliers | $ATE_c$ |
| $T_i \ \vert T_i$ | Always treats | $ATE_a = 0$ |
| $C_i \ \vert C_i$ | Never treats | $ATE_n = 0$ |

$Overall\, ATE = p_c ATE_c + p_a ATE_a + p_n ATE_n = p_c ATE_c$
Therefore $ATE_c = \frac{Overall\, ATE}{p_c}$

In the assigned-to-treatment group:

| | |
|---|---|
| $T_i \ \vert\vert\vert\vert$ | Compliers or Always-treats |
| $\vert\vert\vert\vert\vert C_i$ | Never-treats $\Longleftarrow$ tell people to serve but some don't |

In the assigned-to-control group:

| | |
|---|---|
| $\vert\vert\vert\vert\vert C_i$ | Compliers or Never-treats |
| $T_i \ \vert\vert\vert\vert$ | Always-treats $\Longleftarrow$ tell people **not** to serve but some do serve |

Fraction accept treatment in treatment group: $p_c + p_a$
Fraction accept treatment in control group: $p_a$
Therefore we can get $p_c$ by deducting the second one from the first one: $p_c = (p_c + p_a) - p_a$

- Instrumental Variable
  The effect will break when:
  - Confound variable influences instrumental variable
  - Instrumental variable influences DV

  Another example of instrumental variable:

  $$\overline{\underline{Weather\, in\, city\, A}} \Longrightarrow \overline{\underline{Running\, in\, city\, A}} \Longrightarrow \overline{\underline{Running\, in\, city\, B}}$$

  The instrumental variable (weather in city A) only changes the probability of IV (Running in city A) so that we can figure out if IV **causes** DV.