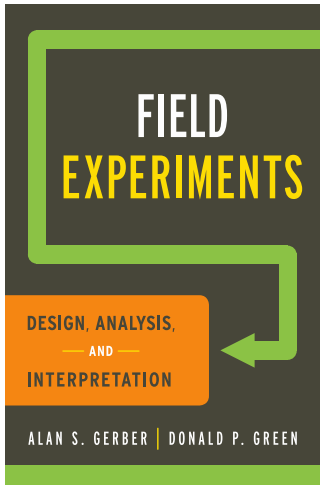
J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

Threats to Inference and Ways to Design Field Experiments to Guard Against Them

Don Green
Department of Political Science,
Columbia University
donald.p.green@gmail.com



Outline

- Core assumptions and implications for:
  - Design
  - Analysis
  - Interpretation

- Threats to Inference:
  - Noncompliance
  - Attrition
  - Spillovers

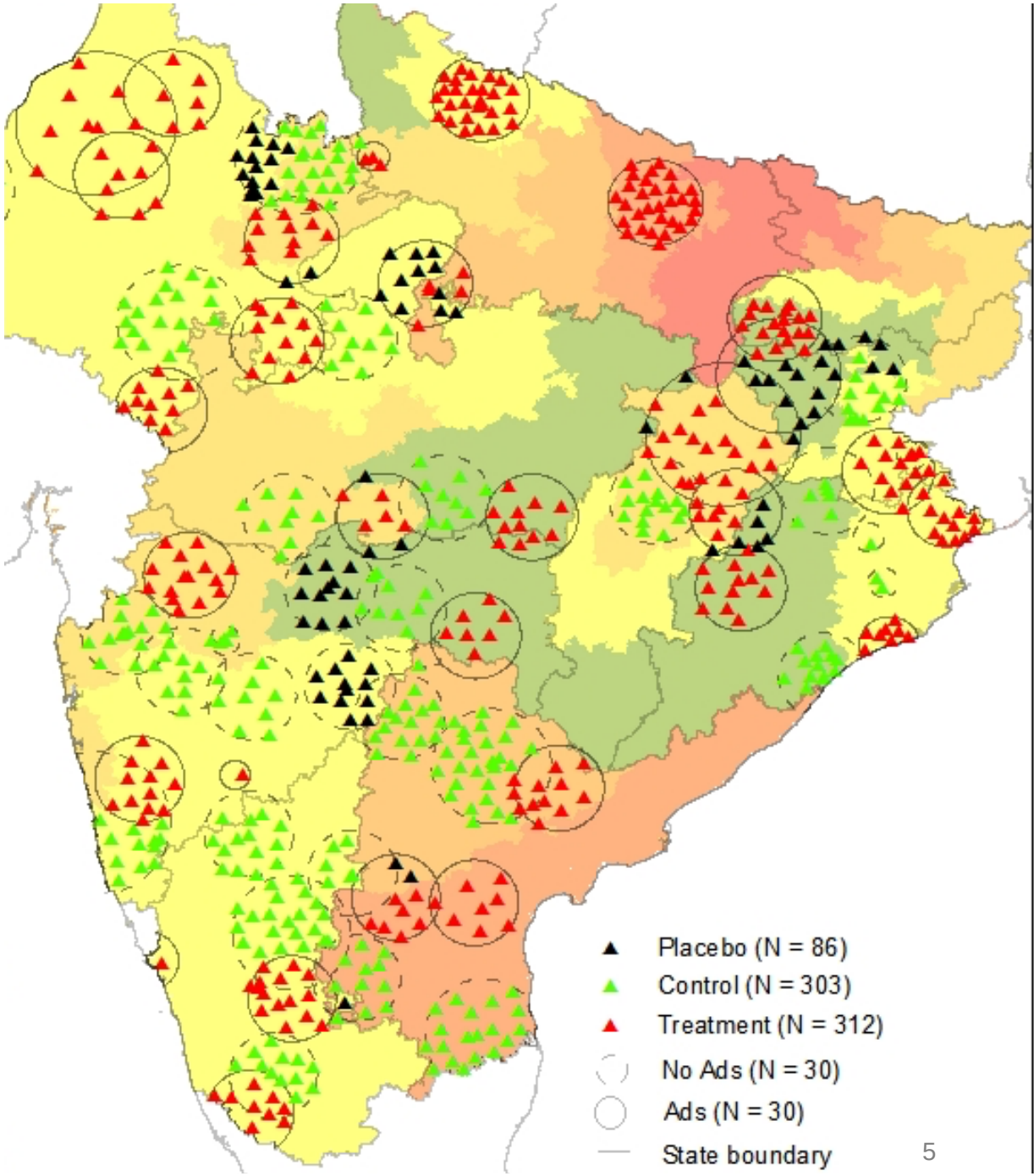
- Generalizability

Brief sketches of some current experimental projects

- India: effect of radio messages to discourage voters from supporting vote-buying parties

- Uganda: video vignettes about domestic violence, abortion, and teacher absenteeism

Assigned treatments by election phase

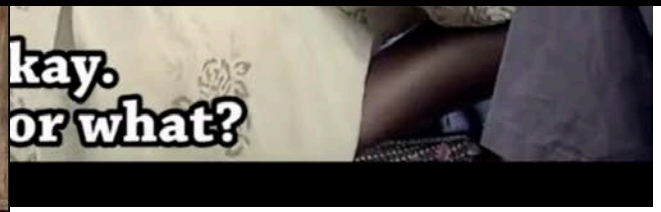


Blocked assignments of Ugandan trading centers to three media messages or combinations of messages

Storylines from the three three-part vignettes

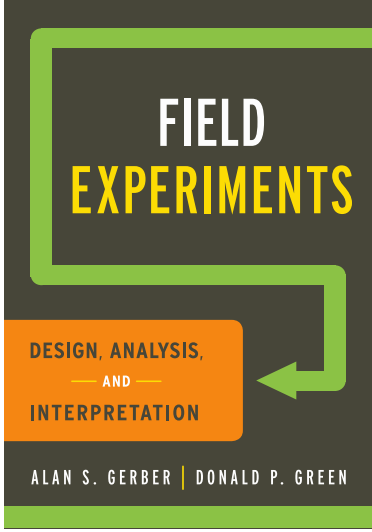
Domestic violence

Abortion



Teacher absenteeism

Themes of FEDAI



Importance of…

- Defining the estimand
- Appreciating core assumptions under which a given experimental design will recover the estimand
- Conducting data analysis in a manner that follows logically from the randomization procedure
- Following procedures that limit the analyst's discretion in data analysis
- Presenting results in a detailed and transparent manner

Potential outcomes and causal effects

- How would an experimental subject have responded if treated?
- How would same subject have responded if untreated?
- Difference between these two potential outcomes is the unit-level treatment effect
- Average unit-level treatment effect is the ATE (average treatment effect in the subject pool)

E.g. (hypothetical) schedule of potential outcomes

Location	Policy support if NOT exposed to media vignettes	Policy support if exposed to media vignettes	Treatment effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
AVERAGE	**15**	**20**	***5***

Core assumptions

- Random assignment of subjects to treatments
  - receiving treatment statistically independent of subjects' potential outcomes

- Non-interference: subject's potential outcomes reflect only whether they receive the treatment themselves
  - Subject's potential outcomes unaffected by how treatments happened to be allocated

- Excludability: subject's potential outcomes respond only to defined treatment, not other extraneous factors that may be correlated with treatment
  - Importance of defining treatment precisely and maintaining symmetry between treatment and control groups (e.g., through blinding)

Conspicuously absent from list of core assumptions…

- No assumptions about shape of outcome distribution (e.g., that responses are normally distributed)

- The issue of "external validity" is a separate question that relates to the issue of whether the results obtained from a given experiment apply to other subjects, treatments, contexts, and outcomes

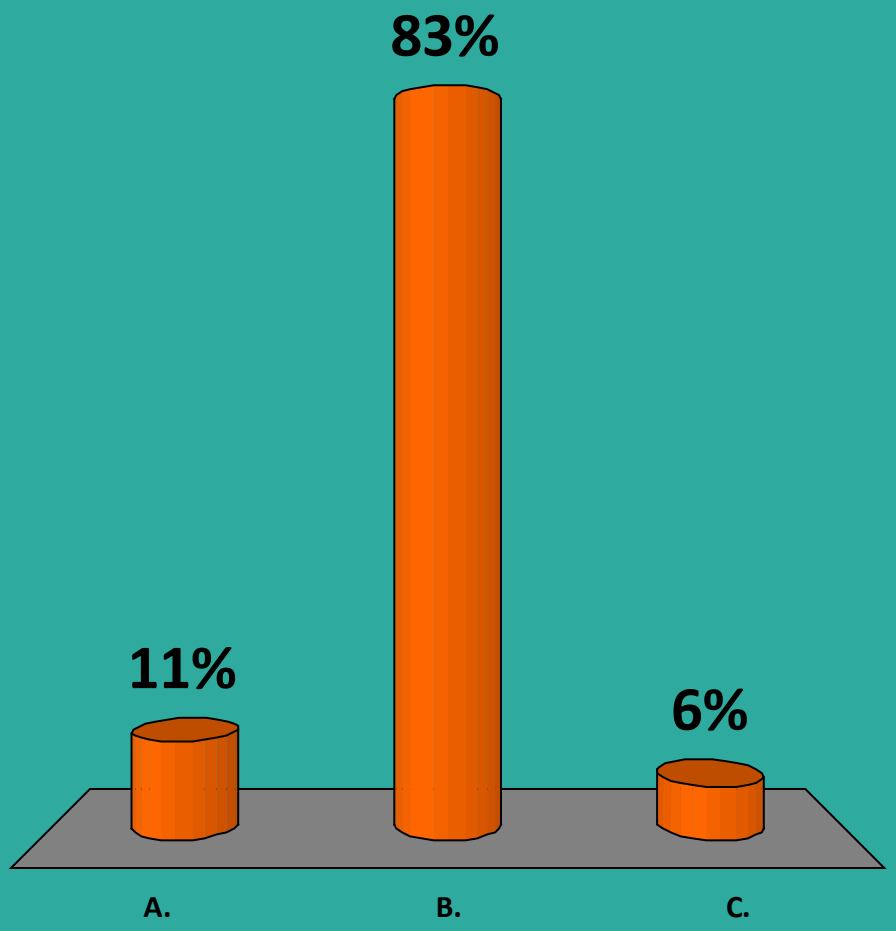
- Random sampling of subjects from a larger population not a core assumption, but aids generalizability

Key result: When core assumptions are met, an experiment generates unbiased estimates of the ATE

- Sampling distribution: collection of possible ways that an experiment could have come out, under different random assignments
- An estimator is a procedure for generating guesses about a quantity of interest (e.g., the average treatment effect)
- Under simple or complete random assignment, the difference-in-means estimator is unbiased
  - Any given estimate may be higher or lower than the true ATE, but on average, this procedure recovers the correct answer

Random assignment will always give you the true average treatment effect

A. True
B. False
C. Don't know



Noncompliance

- Sometimes there is a disjunction between the treatment that is assigned and the treatment that is received
  - Miscommunication and administrative mishaps
  - Subjects may be unreachable
  - Encouragements sometimes don't work

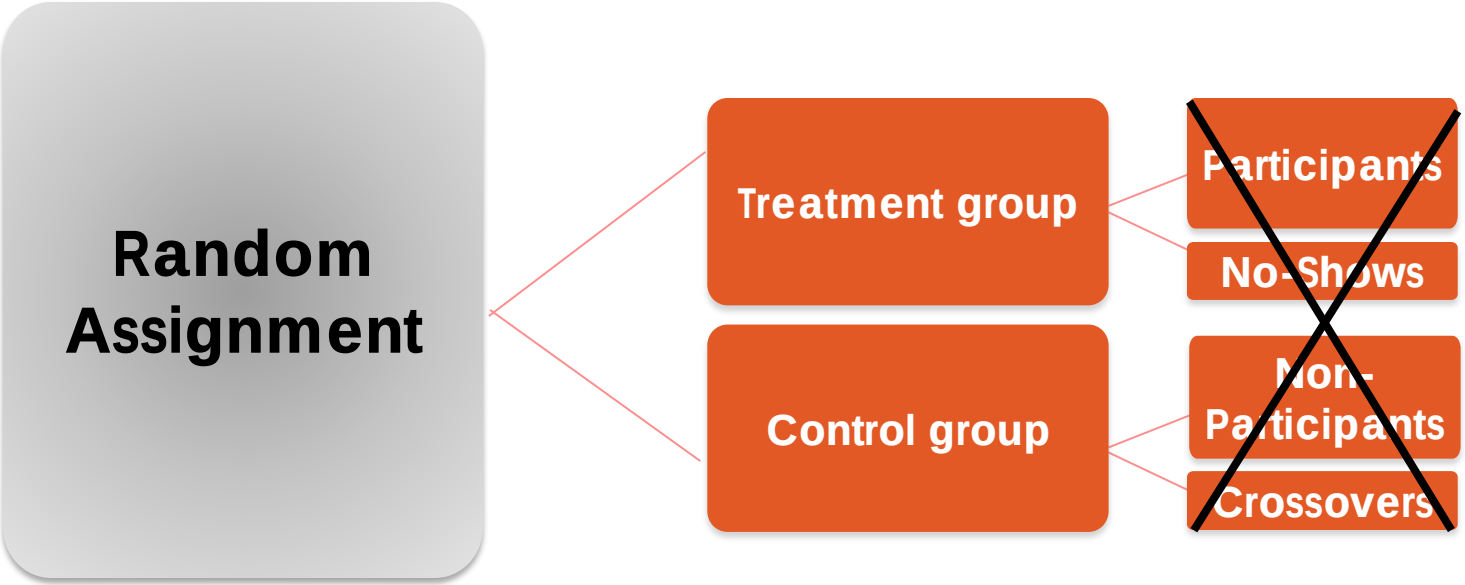
- Addressing noncompliance requires careful attention to "excludability" assumptions
  - Are outcomes affected only by the treatment? Or by both the assignment and the treatment?

Handling noncompliance

What can you do?

Can you switch them?

Bad idea: biased

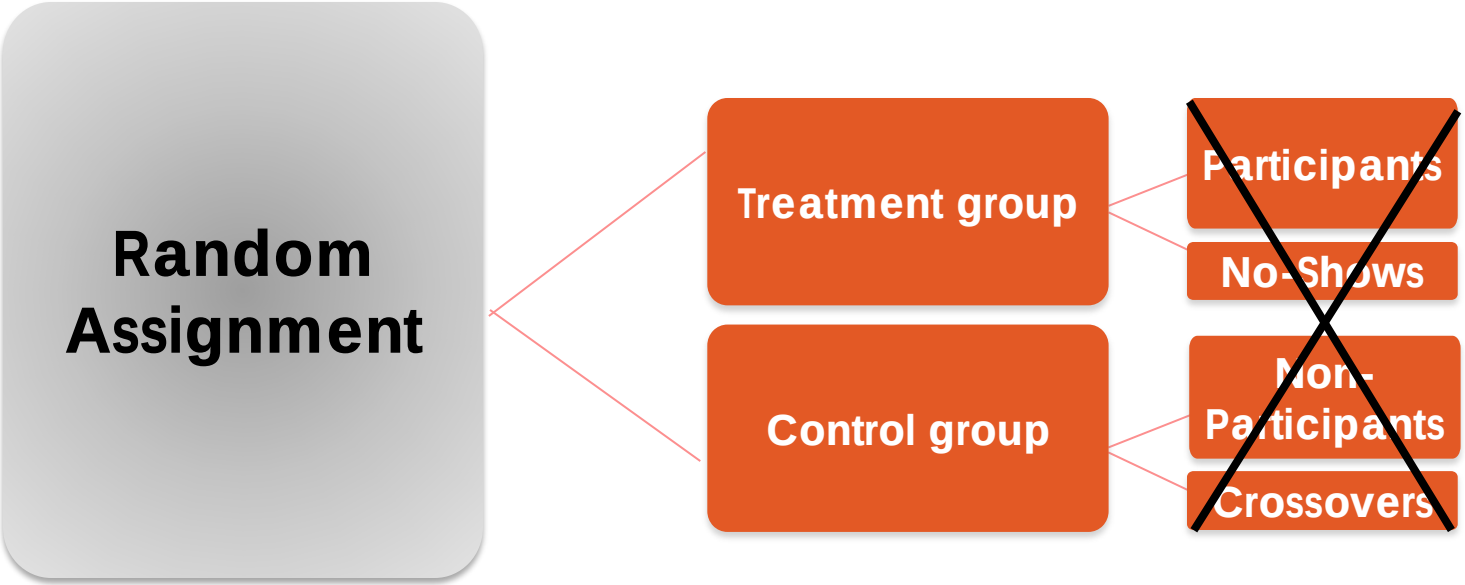


Handling noncompliance

What can you do?

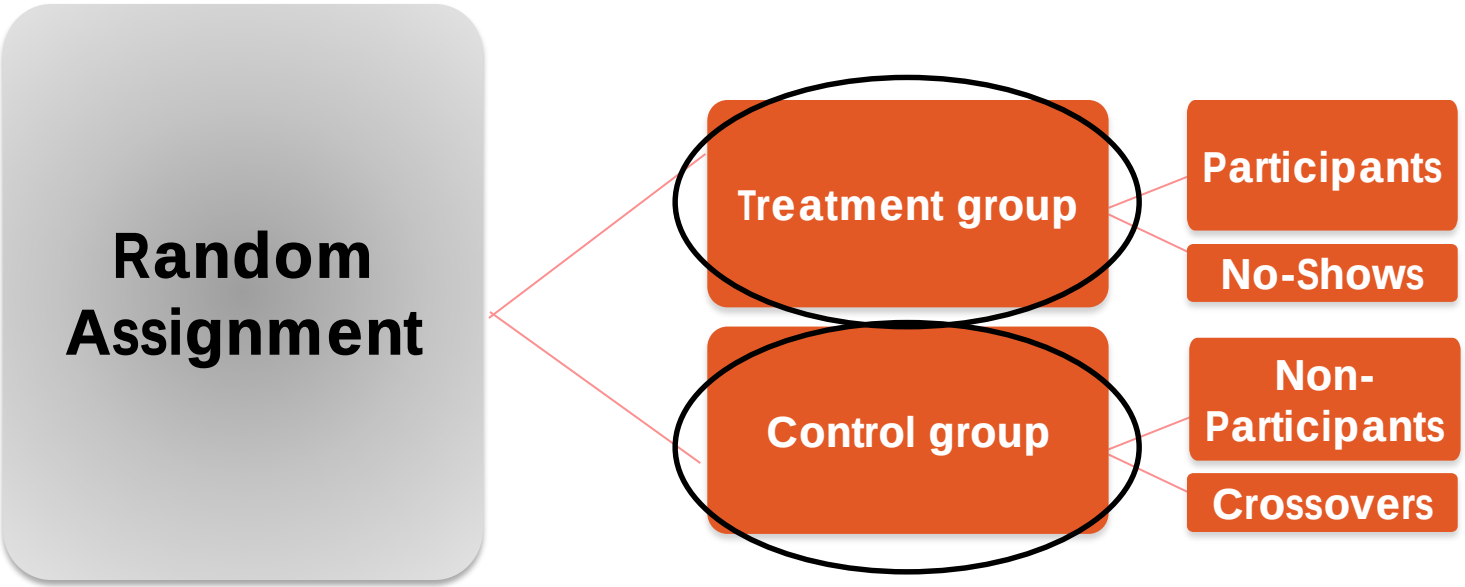
Can you drop them?

Bad idea: biased



Handling noncompliance

Inferences should be based solely on comparisons of randomly assigned groups



Noncompliance: avoiding common errors

- Subjects you fail to treat are NOT part of the control group!
- Do not throw out subjects who fail to comply with their assigned treatment
- Base your estimation strategy on the ORIGINAL treatment and control groups, which were randomly assigned and therefore have comparable potential outcomes

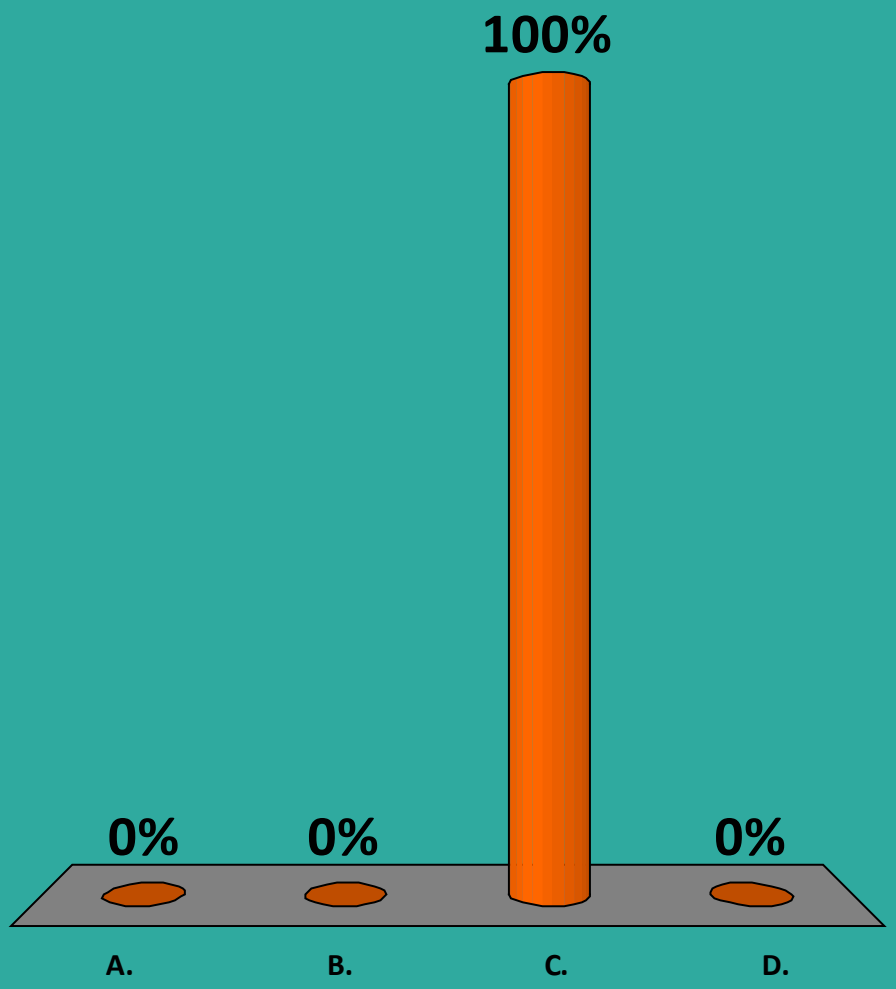
Your treatment group for analysis is…

A. Individuals assigned to treatment who were *actually* treated

B. All individuals who were *actually* treated

C. Individuals assigned to treatment, regardless of whether or not they were treated

D. Don't know



Addressing (one-sided) noncompliance statistically

- Define "Compliers" and estimate the average treatment effect within this subgroup

- Model the expected treatment and control group means as weighted averages of latent groups, "Compliers" and "Never-takers"

- Assume excludability: assignment to treatment only affects outcomes insofar as it affects receipt of the treatment (the plausibility of this assumption varies by application)

Compliers and Never-takers defined

- Compliers: those subjects who would take the treatment if and only if assigned to the treatment group

- Never-takers: those subjects who never receive the treatment regardless of their assignment

- Notice that Compliers and Never-takers are defined in terms of their potential response to assignment

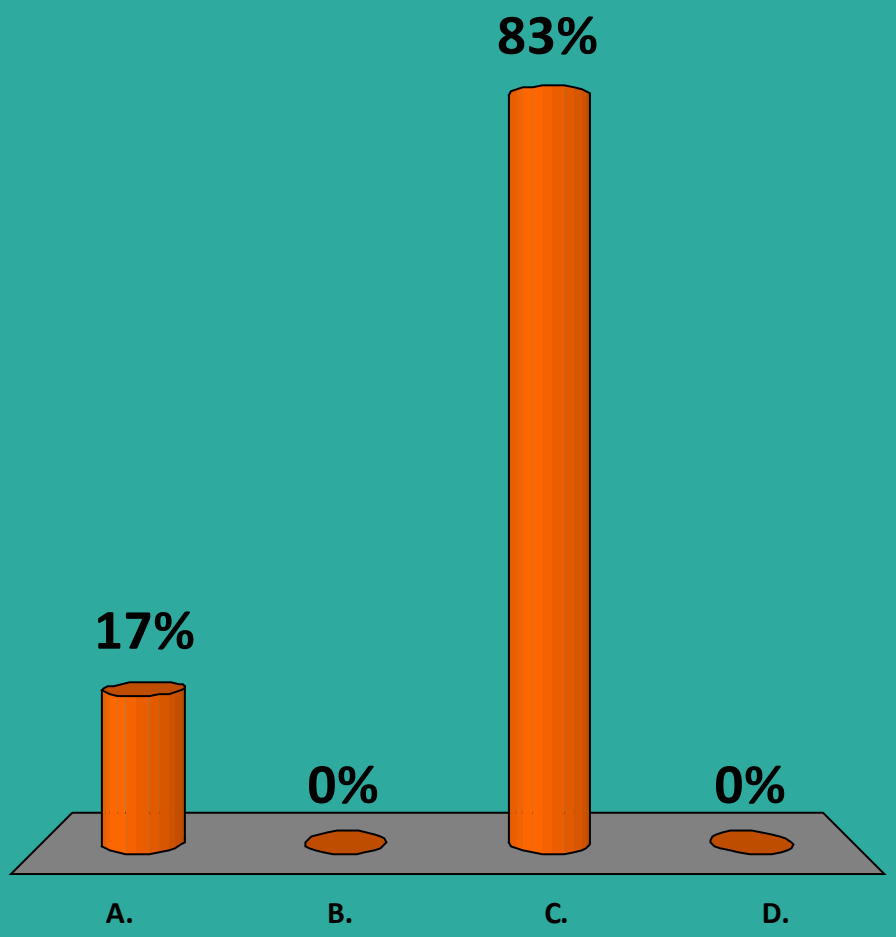
Compliers are:

A. Individuals who *always* take up the treatment

B. Individuals who *never* take up the treatment

C. Individuals who would take up the treatment only if assigned to the treatment group

D. Don't know



Example: The effects of viewing a film about abortion stigma in rural Uganda

- In 2015, 56 villages were randomly assigned to air a series of films; in some are embedded videos about helping those who suffer from medical complications arising from illegal abortions
- Intended treatment is exposure to abortion messages, but only some villagers actually turn up to watch the film festival
- "Compliers" are those who would be exposed to the abortion treatment if assigned to it
- "Never-takers" are those who would not be exposed to the abortion treatment regardless of whether they are assigned to it
- Outcome: expressing empathy for those ostracized because of abortion in the end-line survey

Simplified model notation

- Suppose that the subject pool consists of two kinds of people: Compliers and Never-takers
- Let P_c = the probability that untreated Compliers express empathy for those ostracized because of abortion in the end-line survey
- Let P_n = the probability that untreated Never-takers express empathy for those ostracized because of abortion in the end-line survey
- Let a = the proportion of Compliers in the subject pool
- Let T = the average treatment effect among Compliers

Expected outcomes in control and treatment groups

Expected outcome in the control group (E_0) is a weighted average of Complier and Never-taker outcomes in the control condition:

$$E_0 = \alpha P_c + (1 - \alpha) P_n$$

Expected outcome in the treatment group (E_1) is also a weighted average of Complier and Never-taker outcomes, this time under treatment:

$$E_1 = \alpha (P_c + T) + (1 - \alpha) P_n$$

Derive an Estimator for the Average Treatment Effect among Compliers (T)

$$E_1 - E_0 = a(P_c + T) + (1 - a)P_n - \{a P_c + (1 - a) P_n\}$$
$$= aT$$

aT is the "intent to treat" effect, or "ITT"

To estimate T, insert sample values into the formula:

$$T^* = (E^*_1 - E^*_0)/a^*$$

where a^* is the proportion of treated people (Compliers) observed in the assigned treatment group, and E^*_1 and E^*_0 are the observed average outcomes in the assigned treatment and control groups, respectively

Terminology footnote: LATE = CACE

- Local average treatment effect (LATE) is the same as the Complier Average Causal Effect (CACE)

- Also, note that among Compliers, the CACE = ITT

- Finally, note that among Never-takers, the ITT = 0

Example: Uganda social norms experiment

- Assignment at the trading center level (N=56) to receive videos on abortion or not during commercial breaks in a film festival
- Define Z=1 as the assignment to treatment (abortion message); Z=0 is no abortion message
- Define D=1 as exposure to the abortion message (attended the film or had a friend/relative who attended); D=0, otherwise
- Define Y=1 as willingness to help a girl who was ostracized on account of having an abortion, Y=0 otherwise

Outcome measure: empathy for someone stigmatized on account of abortion

Suppose that a girl in your neighborhood has had a deliberate abortion:

- because she wanted to stay in school
- because she wanted to take a full-time job

She has been ostracized from the community and people seem to have turned their backs on her.) In this situation, two of your friends make the following two statements. Which friend would you agree with?

- She made her choice and has violated God's rule, it is better not to get involved. (Coded 0)
- Regardless of what this woman did, we should try to help her. (Coded 1)

Estimate the effect of assignment (Z) on treatment (D): a* =0.682

D	Z: Control	Z: ABO Treat	Total
Untreated	1,390	331	1,721
	100.00	31.80	70.79
Treated	0	710	710
	0.00	68.20	29.21
Total	1,390	1,041	2,431
	100.00	100.00	100.00

Estimate E^*_1=0.7560, E^*_0=0.7079

Notice that estimated ITT is 0.0481

Y	Z: Control	Z: ABO Treat	Total
No Help	406	254	660
	29.21	24.40	27.15
Help	984	787	1,771
	70.79	75.60	72.85
Total	1,390	1,041	2,431
	100.00	100.00	100.00

Estimate Complier Average Causal Effect

$(E^*_1 - E^*_0) / a^* = (0.7560 - 0.7079) / 0.6820 = 0.0705$

In other words: among Compliers, actual exposure increased the probability of expressing empathy by 7.1 percentage-points

This estimator is equivalent to instrumental variables regression, where assignment to treatment is the instrument for actual exposure

Notice that we NEVER compare the outcomes of those who were exposed to the treatment to the outcomes among those who were not exposed…Why not?

Equivalent estimates from OLS (ITT)

```
. reg y z , cl(TradingCenter)

Linear regression                               Number of obs =    2431
                                                F(  1,    55) =    4.79
                                                Prob > F      =  0.0329
                                                R-squared     =  0.0029
                                                Root MSE      =  .44428

                          (Std. Err. adjusted for 56 clusters in TradingCenter)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           z |   .0480902   .0219703     2.19   0.033     .0040607    .0921197
       _cons |   .7079137   .0175496    40.34   0.000     .6727435    .7430838
------------------------------------------------------------------------------
```

Equivalent estimates from Instrumental Variables Regression (CACE)

```
. ivregress 2sls y (d = z), cl(TradingCenter)

Instrumental variables (2SLS) regression          Number of obs =    2431
                                                  Wald chi2(1)  =    4.83
                                                  Prob > chi2   =  0.0279
                                                  R-squared     =  0.0015
                                                  Root MSE      =   .4444

                        (Std. Err. adjusted for 56 clusters in TradingCenter)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           d |   .0705097   .0320671     2.20   0.028     .0076593    .1333601
       _cons |   .7079137   .0173886    40.71   0.000     .6738326    .7419947
------------------------------------------------------------------------------
```

Alternative design: Placebo control

- Placebo must be (1) ineffective and (2) administered in the same way as the treatment, such that assignment to placebo/treatment is random *among those who are exposed to an intervention*
  - Assignment to placebo should be blinded and made at the last possible moment before treatment

- Placebo design can generate more precise estimates, especially when treatment rates are low

Leveraging the Placebo Design in the Uganda Study to estimate the CACE Using a Different Estimator

- Uganda study measured whether respondents attended the festival (full compliers), had friends/relatives who attended, or neither

- Compliance types are revealed in both treatment and control groups due to the placebo design (assuming symmetry!)

- We find no effect among Never-takers, consistent with assumptions of the design

Compare estimated treatment effects among Compliers (those who attended the film or had friends/relatives who attended)

Y	Z: Control	Z: ABO Treat	Total
No Help	112	84	196
	25.34	25.38	25.36
Help	330	247	577
	74.66	74.62	74.64
Total	442	331	773
	100.00	100.00	100.00

Estimated effect among Never-takers should be close to zero – and is

Y	Z: Control	Z: ABO Treat	Total
No Help	294	170	464
	31.01	23.94	27.99
Help	654	540	1,194
	68.99	76.06	72.01
Total	948	710	1,658
	100.00	100.00	100.00

Estimated effect among Compliers is 7.07

Equivalent CACE estimate from OLS

```
Linear regression                               Number of obs =    1658
                                                F(  1,    55) =   11.06
                                                Prob > F      =  0.0016
                                                R-squared     =  0.0061
                                                Root MSE      =  .44783

                          (Std. Err. adjusted for 56 clusters in TradingCenter)
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           z |     .07069   .0212525     3.33   0.002      .028099    .1132809
       _cons |   .6898734   .0159036    43.38   0.000     .6580018     .721745
------------------------------------------------------------------------------
```

Notice the standard error is smaller than above, illustrating the greater precision of the placebo design

Summary: CACE and ITT

- Failure to treat changes your estimand and diminishes power
- The CACE is the "Complier Average Causal Effect"
  - Compliers = those who take the treatment if and only if assigned to the treatment
- The CACE may be quite different from the ATE, so be careful about extrapolation: Compliers may be different than those forced
- ITT is the average effect of assignment, which is useful for program evaluation, because failure-to-treat is a relevant feature of a program

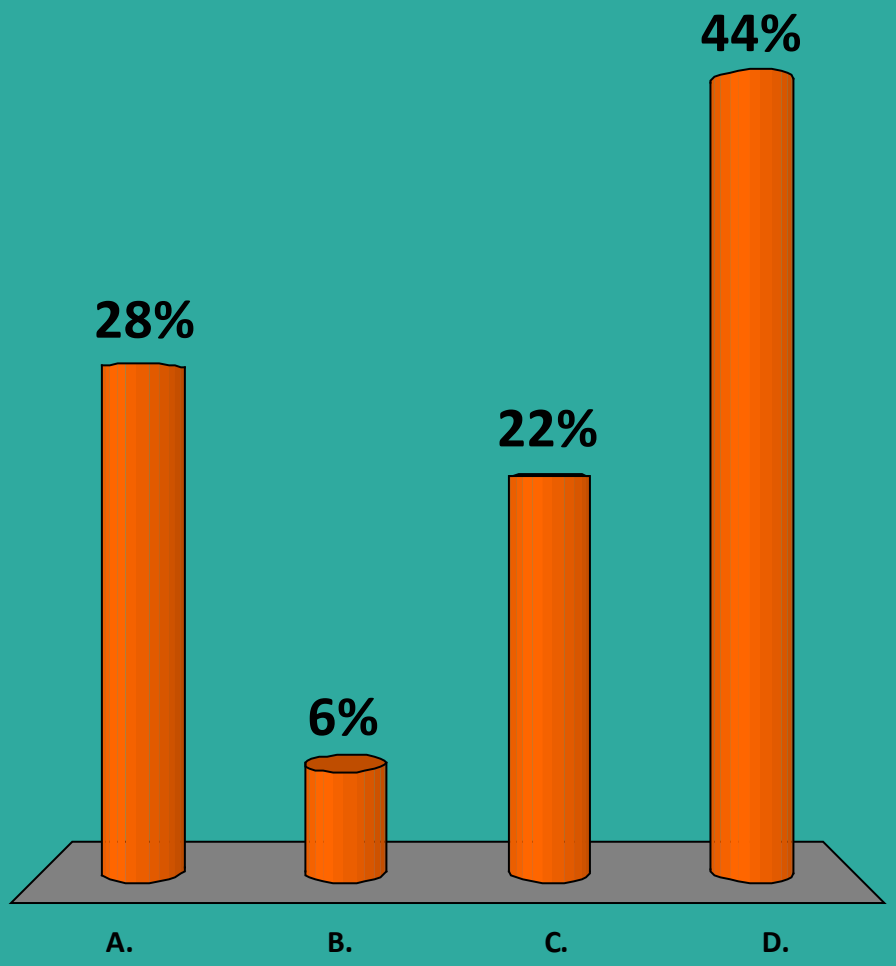
The $\widehat{CACE}$ and the $\widehat{ITT}$ effect will be the same…

A. …when there is perfect compliance

B. …when the ATE is 0

C. …amongst compliers

D. Answers A and C



Remember the assumptions behind estimation of the CACE

- Assignment has no "backdoor path" to outcomes: (Example of an application where this assumption is problematic: restorative justice)

- You must define the treatment and measure whether subjects receive the treatment

- In the simple example used here, the control group is never treated inadvertently. Two-sided noncompliance is somewhat more complicated, but frequently arises with "encouragement designs"

Review

- What is the difference between the ITT and the ATE?
- Define "Complier."
- Why is it incorrect to say "never-takers are those who are untreated"?
- How does one estimate the proportion of compliers in the subject pool?
- What is a Complier Average Causal Effect?
- True or false: "the ITT always has the same sign as the CACE"
- True or false: "A placebo-controlled design provides an unbiased estimate of the CACE provided that the rate of compliance in the placebo group is the same as the rate of compliance in the treatment group."
- Suppose the mean outcome in the assigned control group is 35. The mean in the assigned treatment group is 45. In the assigned treatment group, 25% of the subjects receive the treatment; no one is treated in the assigned control group. The mean outcome among those actually treated is 35. Estimate the CACE.

Attrition

- Can present a grave threat to any experiment because missing outcomes effectively "un-randomize" the assignment of subjects

- Example: Rand Health Insurance Experiment
  - Offered subjects chance to receive either 5% or 100% health coverage (or intermediate coverage levels)
  - Differential rates of refusal and dropout among those offered low levels of coverage
  - Threat of bias: What if those who expected to become sick dropped out when offered low levels of health coverage?

Diagnosing the Risk of Attrition-related Bias

If you confront attrition, consider whether it threatens the symmetry between assigned experimental groups

- Are the rates of attrition the same in the treatment and control groups?
- Do covariates predict missingness in the same way in both the treatment and control groups?

Don’t introduce attrition unwittingly

- Bad: When analyzing a lab experiment, throwing out subjects in the treatment group who, in a debriefing interview after the experiment concludes, indicate that they had figured out what the experimental hypothesis was

- Bad: When analyzing donations, throwing out those who don’t contribute anything and estimating the ATE based solely on those who contribute a positive amount

- Maybe OK: Throwing out those in the treatment and control group for whom administrative data were unavailable for reasons that seem unrelated to treatment

Attrition: Design-based solutions…and non-solutions

- Consider design-based solutions such as an intensive effort to gather outcomes from a random sample of the missing (“double sampling”)
- Imputation of “extreme values” to obtain worst-case bounds on the ATE
- Use of “trimming” bounds plus a “monotonicity” assumption in order to bound the ATE among those who would always provide outcomes regardless of treatment assignment
- Beware of dropping missing observations or blocks in which missingness occurs

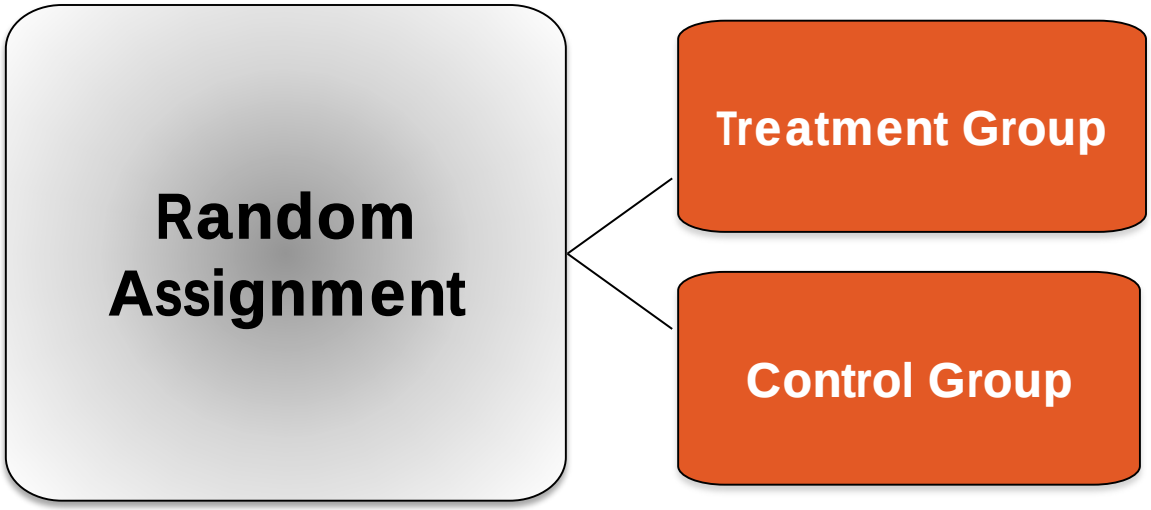
Missing outcome measures vs. missing covariates

- Missing outcomes: potential for bias

- Missing covariates: since covariates are optional, it is not advisable to drop observations with missing covariates

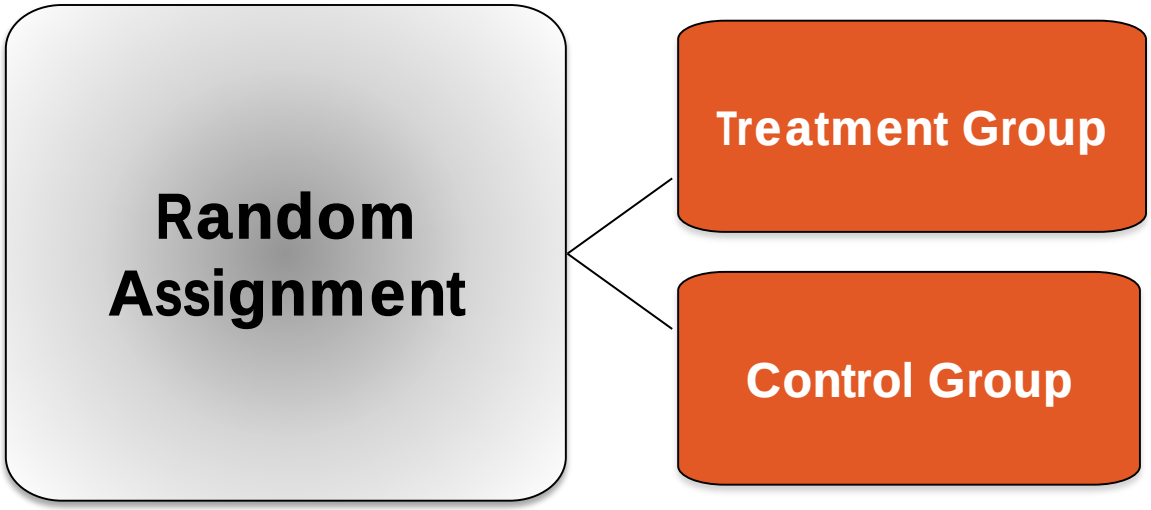
- In order to keep all subjects in the analysis…
  - impute missing values for a covariate or
  - in the context of regression analysis, insert an arbitrary value for the covariate and include a dummy variable that indicates missingness

Another potential problem: interference



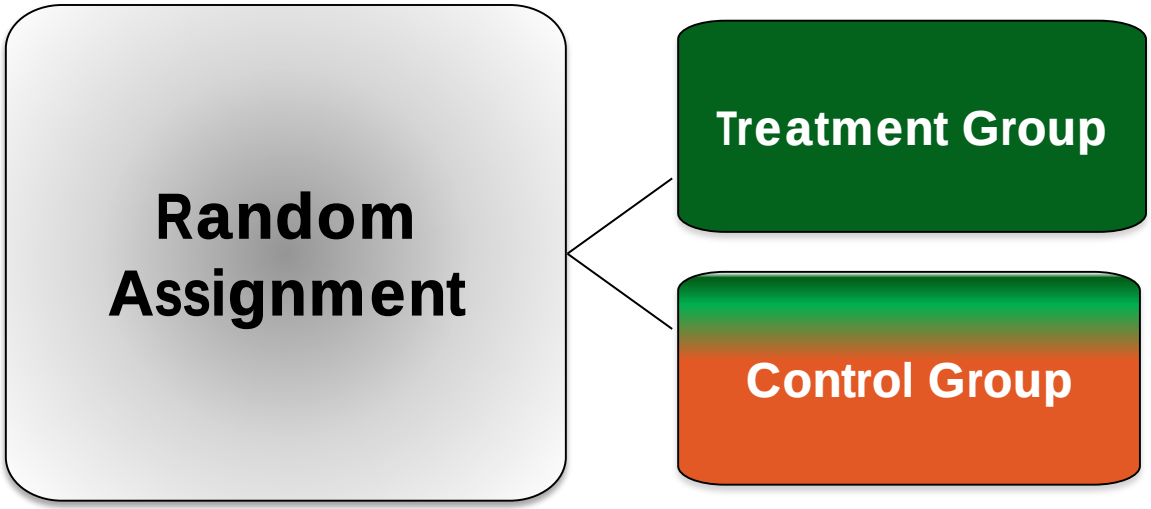
What else could go wrong?

Treatment applied....



Spillovers…

Treatment applied….



Spillovers

- Are subjects' potential outcomes a reflection ONLY of whether or not they personally receive treatment?
- Or could it be that subjects are affected as well by which other subjects receive treatment?

Hypotheses about spillovers

Contagion: The effect of being vaccinated on one's probability of contracting a disease depends on whether others have been vaccinated.

Displacement: Police interventions designed to suppress crime in one location may displace criminal activity to nearby locations.

Communication: Interventions that convey information about commercial products, entertainment, or political causes may spread from individuals who receive the treatment to others who are nominally untreated.

Social comparison: An intervention that offers housing assistance to a treatment group may change the way in which those in the control group evaluate their own housing conditions.

Persistence and memory: Within-subjects experiments, in which outcomes for a given unit are tracked over time, may involve "carryover" or "anticipation."

Spillovers greatly complicate statistical analysis

- Complication: equal-probability random assignment of units does not imply equal-probability assignment of exposure to spillovers

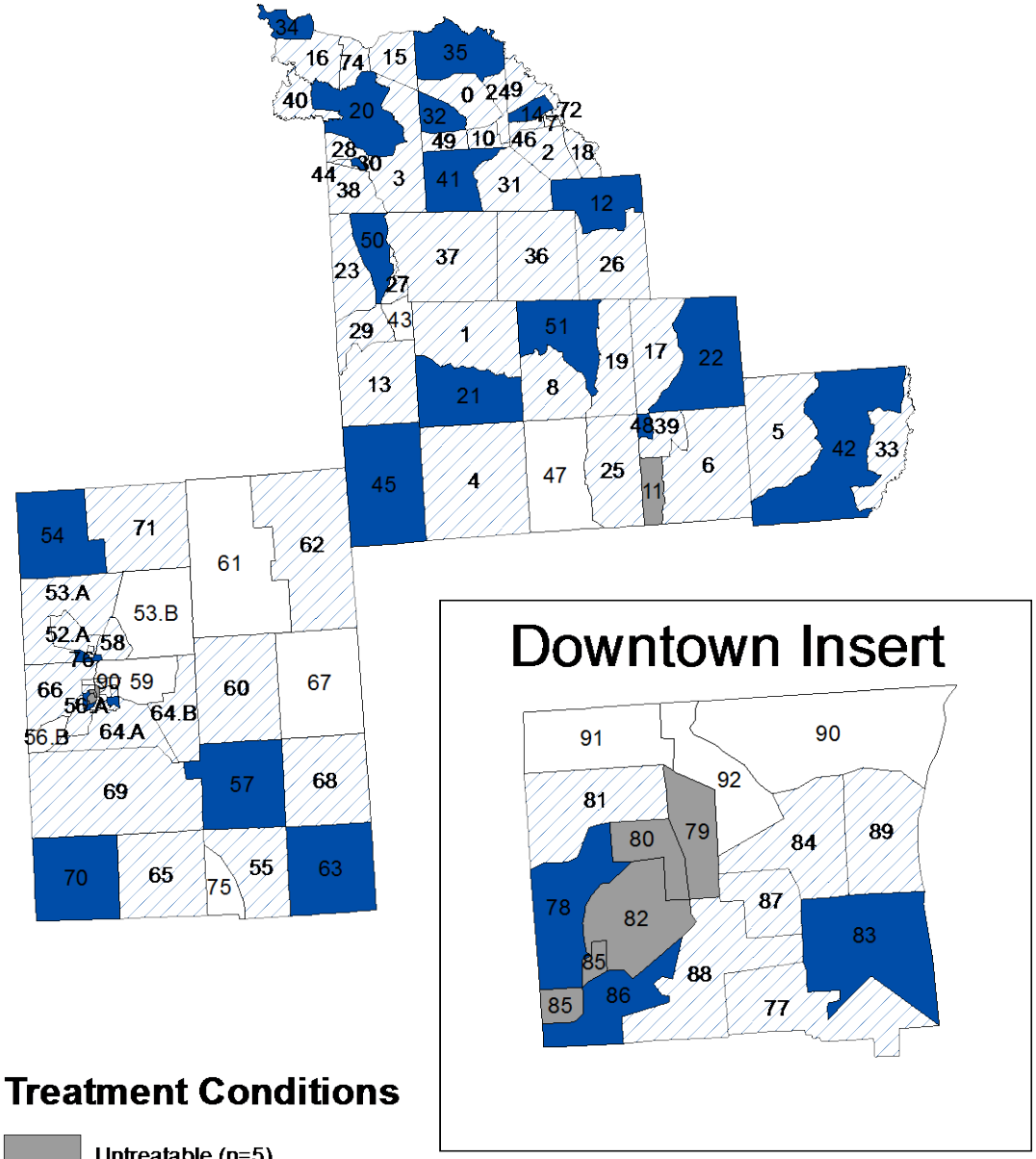
- Unweighted difference-in-means (or unweighted regression) can give severely biased estimates

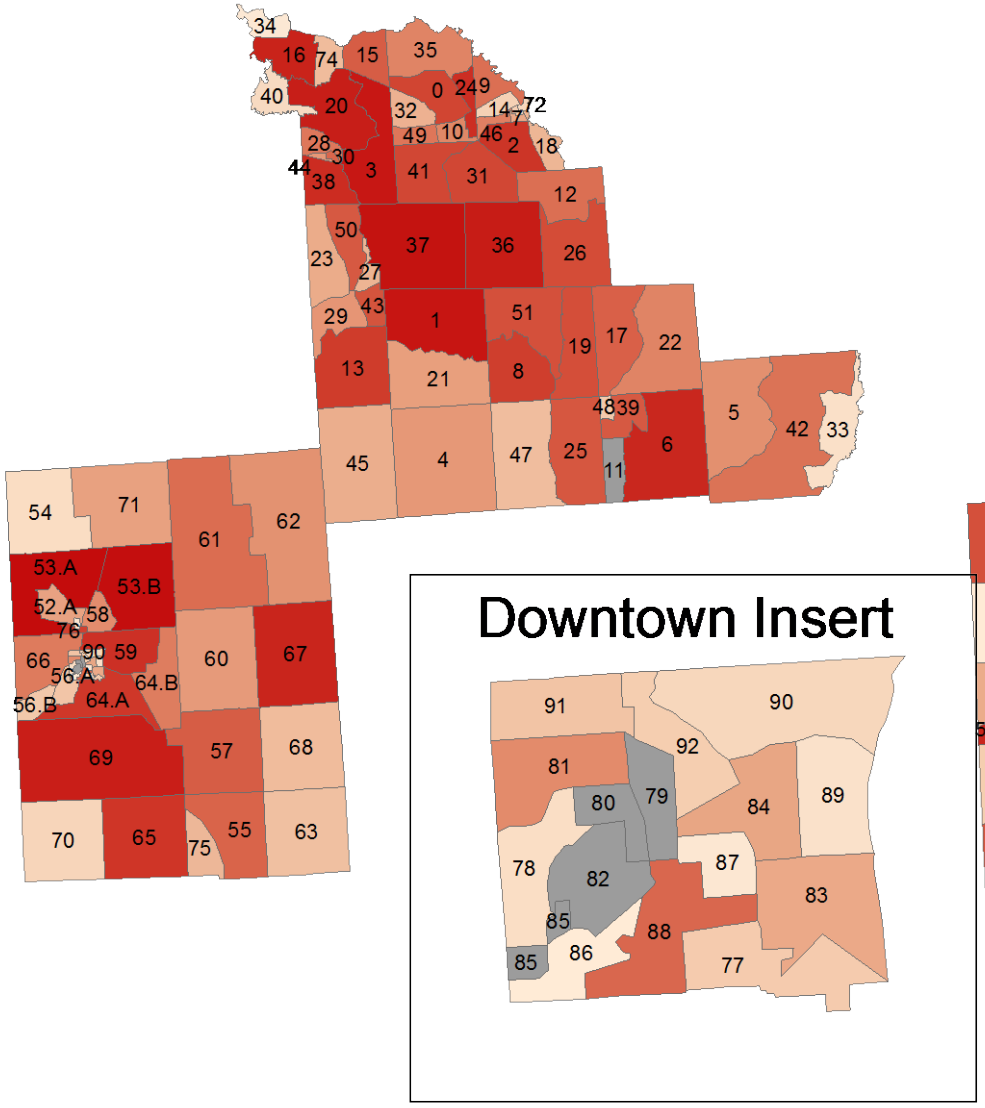
Example: Assessing the effects of lawn signs on a congressional candidate's vote margin

- Complication: the precinct in which a lawn sign is planted may not be the precinct in which those who see the lawn sign cast their votes

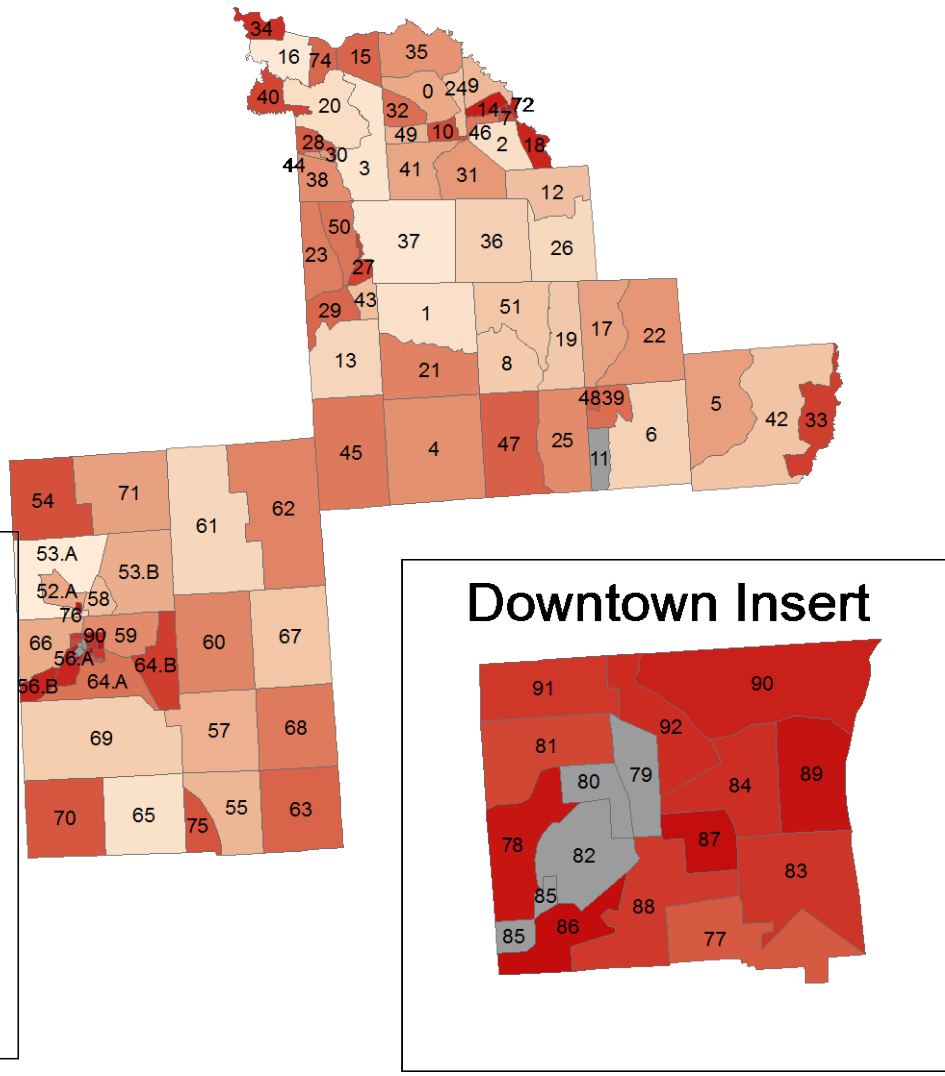
- Exposure model: define a potential outcome for (1) precincts that receive signs, (2) precincts that are adjacent to precincts with signs, and (3) precincts that are neither treated nor adjacent to treated precincts

- Further complication: precincts have different probabilities of assignment to the three conditions


Downtown Insert
Treatment Conditions
Untreatable (n=5)
Directly Treated (n=23)
Adjacent to Treated (n=57)
Control (n=12)



Deeper Reds Indicate Higher Probability of Assignment to Spillovers (0,1)

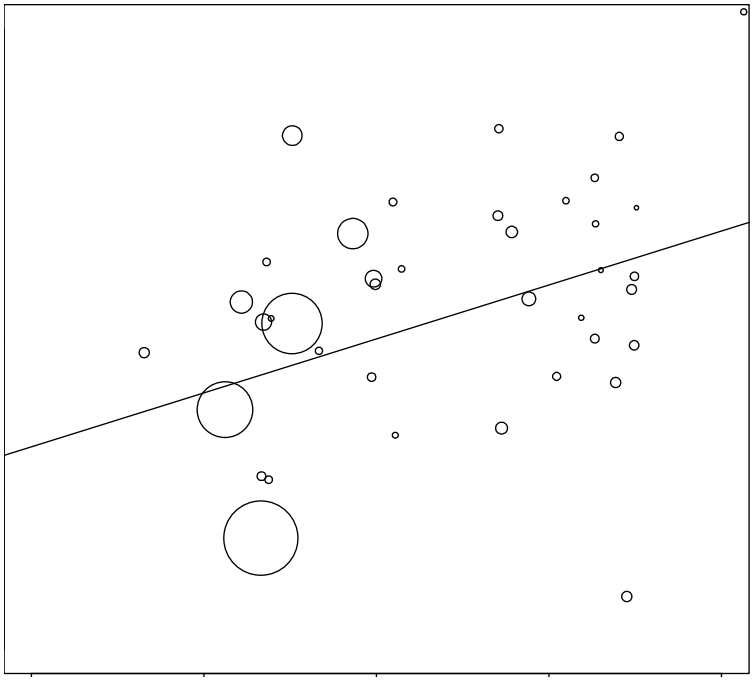


Deeper Reds Indicate Higher Probability of Assignment to Control (0,0)

Table 2: Weighted Least Squares Estimates of the Effects of Lawn Signs on Congressional Vote Margin and Vote Share

	Congressional Vote Margin		Congressional Vote Share	
Assigned Lawn Signs (n=23)	15.582	34.786	0.025	0.025
	(30.690)	(20.393)	(0.029)	(0.020)
Adjacent to Lawn Signs (n=49)	−9.242	11.713	0.037	0.018
	(29.621)	(20.046)	(0.029)	(0.018)
Covariate Adjustment	no	yes	no	yes
N	88	88	88	88
R^2	0.015	0.666	0.031	0.823

Column 2 covariates: Congressional Vote Margin 2006-10 and Presidential Vote Margin 2008.
Column 4 covariates: Congressional Vote Share 2006-10 and Presidential Vote Share 2008.
Robust standard errors in parentheses.



st Indirect Treatment

Residualized Vote Margin

Residualized indirect Treatment

Non-interference: Summing up

- Unless you specifically aim to study spillover, displacement, or contagion, design your study to minimize interference between subjects. Segregate your subjects temporally or spatially so that the assignment or treatment of one subject has no effect on another subject's potential outcomes.

- If you seek to estimate spillover effects, remember that you may need to use inverse probability weights to obtain consistent estimates

Drawing generalizations?

- Four considerations regarding generalizability
  - Subjects
  - Treatments
  - Contexts
  - Outcomes

Design with generalization in mind

- The immediate aim is to estimate the ATE in our subject pool
  - Eventually, we may seek to replicate our study using different subjects, treatments, contexts, or outcomes and/or to explore systematic sources of treatment effect heterogeneity

- One should look for opportunities to conduct a series of experiments on the same subject pool (e.g., by using the control group of study 1 as the subject pool for study 2)

- Also advisable to include the same treatment when studying different subject pools to calibrate effects

Extra slides – additional topics

Identification of CACE for two-sided noncompliance: Mullainathan, Washington, and Azari 2010 study of NYC mayoral debates

- Partition the sample into Always-takers, Compliers, and Never-takers (assume there are no Defiers)
- Let P_A = expected outcomes for untreated Always-takers
- Let P_C = expected outcomes for untreated Compliers
- Let P_N = expected outcomes for untreated Never-takers
- Let a_A = the proportion of subjects who are Always-takers
- Let a_C = the proportion of subjects who are Compliers (the remainder are Never-takers)
- Let T_C = the ATE among Compliers (adjust subscripts for Always-Takers and Never-Takers)

Estimating the CACE under two-sided noncompliance

Expected outcomes in the control group are a weighted average of outcomes for Always-takers, Compliers, and Never-takers (assuming no Defiers):

$$V_0 = a_A(P_A + T_A) + a_C(P_C) + (1-a_A-a_C)\, P_N$$

$$V_1 = a_A(P_A + T_A) + a_C(P_C + T_C) + (1-a_A-a_C)\, P_N$$

Therefore, $V_1 - V_0 = a_C T_C$

Estimating the CACE under two-sided noncompliance (continued)

$V_1 - V_0 = a_C T_C$

The expected fraction of subjects who receive the treatment in the assigned control group is a_A

The expected fraction of subjects who receive the treatment in the assigned treatment group is $a_A + a_C$

Therefore, the expected difference is a_C

To estimate the CACE under two-sided noncompliance, use sample estimates of the following quantities:

$CACE = (V_t - V_c)/(a_A + a_C - a_A) = T_C$

Example of two-sided noncompliance: Mullainathan, Washington, and Azari 2010 study of NYC mayoral debates

$CACE = (V_1 - V_0)/(\alpha_A + \alpha_C - \alpha_A) = T_C$

Sample estimates:

(47.5 – 41.8)/(0.366 – 0.162) = 27.9 ppts

	Treatment group	Control group
% Reporting Watching Debate	36.6	16.2
(N Treated)	(185)	(80)
% Reporting Change Opinions	47.5	41.8
(Total N)	(505)	(495)