Experimental Workshop: Lecture 3

Non-Compliance

Raymond Duch

May 10, 2022

University of Oxford DPIR

Road Map

- Noncompliance

Noncompliance

Intuition

- If $ATE = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$, how can we ever know that subjects were actually treated?
- More importantly, *what does it mean to be "treated"?*
- Let us distinguish between
  - Assignment of treatment (Z)
  - Receipt of treatment (D)
- Yes, by the exclusion restriction, $Y_i(z, d) = Y_i(d)$
- However, in many applications, $z_i \neq d_i$
- Noncompliance with treatment assignment = subjects do not receive the treatment to which they were assigned

Kalla and Broockman 2020

- Kalla and Broockman (2020): Reducing exclusionary attitudes through interpersonal conversation (APSR)
- 230 canvassers *are assigned to* have face-to-face conversations with 6,869 voters deploying non-judgmental exchange of narratives on a range of topics
- Outcome: Exclusionary immigration policy and prejudicial attitudes
- What can go wrong?
- In this example, when does a subject comply with the treatment assignment?

TABLE 1. Summary of Differences Between Conditions and Results in Previous Study and Experiments 1–3

Study	Broockman and Kalla (2016)	Experiment 1		Experiment 2		Experiment 3
Topic	Transphobia	Unauthorized immigrants		Transphobia		Transphobia
Condition name	Full Intervention	Full Intervention	Abbreviated Intervention	Participants' and Video Narratives	Video Narratives Only	Participants' Narratives by Phone
Intervention contents						
Non-judgmental exchange of narratives…						
○ From participants (voter and canvasser)	YES	YES	NO	YES	NO	YES
○ In video	YES	NO	NO	YES	YES	NO
Address concerns and deliver talking points	YES	YES	YES	YES	YES	YES
Results						
ITT[a]	Positive effects ($d = 0.16$, $p < 0.001$)	Positive effects ($d = 0.08$, $p < 0.001$)	Null effects ($d = 0.02$, $p = 0.27$), statistically distinguishable from Full Intervention ($d = 0.06$, $p < 0.01$)	Positive effects ($d = 0.08$, $p < 0.001$)	Positive effects ($d = 0.08$, $p < 0.001$)	Positive effects ($d = 0.04$, $p < 0.001$)
CACE[b]	$d = 0.22$	$d = 0.12$	$d = 0.03$ (Abbreviated vs. Placebo)	$d = 0.10$	$d = 0.10$	$d = 0.08$

Notes: Each Experiment also contained a Placebo condition not shown in the table. These Placebo conditions contained no persuasive content on the topics but are used as a baseline for comparison when estimating the effect sizes shown in the table.

[a]To summarize the results of each study, we first average the pre-specified Overall Index in each study across survey waves to compute a pooled Overall Index. We then report intent-to-treat (ITT) effects on this pooled Overall Index, which represents the mean difference between individuals assigned to each condition among all individuals who identified themselves at their doors, regardless of whether the conversation continued after that point. The ITT estimates represent the average causal effect of attempting to treat people who open their doors, even if they refuse to converse soon after. This means the ITT estimates are "diluted" by the presence of individuals who open the door but do not enter into the conversation.

[b]To estimate the implied Complier Average Causal effect (CACE), or the effect among those who received the intervention, we estimate compliance under a conservative definition of compliance, whether participants got to the "first rating" part of the conversation where they initially told canvassers how they felt about the policy. The CACE estimates represent the average causal effect of treating the people who do

Definition and formalization

- Where is the *ATE* row? What are *ITT* and *CACE*?
- Let $d_i(z)$ denote whether subject i is actually treated when treatment assignment is z
- There are different types of compliance and noncompliance with the treatment
- **Compliers**: $d_i(1) = 1$ and $d_i(0) = 0$ or $d_i(1) > d_i(0)$
- **Never-Takers**: $d_i(1) = 0$ and $d_i(0) = 0$
- **Always-Takers**: $d_i(1) = 1$ and $d_i(0) = 1$
- **Defiers**: $d_i(1) = 0$ and $d_i(0) = 1$ or $d_i(1) < d_i(0)$

Definition and formalization

- These groups are formed *after* random assignment, not formed *by* random assignment $\rightarrow$ they might differ systematically in ways that bias *ATE* estimator
- 2 types of noncompliance
  - One-sided: $d_i(1) = 0$ for some i but $d_i(0) = 0 \; \forall i$ (only compliers and never-takers)
  - Two-sided: additionally, $d_i(0) = 1$ for some i (these can be defiers or always-takers)
- In any experiment facing noncompliance, which subjects *could* make up the treatment group, and which the control group? How might that look like in Kalla and Broockman (2020)?
- What is the problem of naively comparing treated and untreated subjects, i.e. estimate *ATE*?

Estimation of treatment effects under noncompliance

- What groups *could* we compare to unbiasedly estimate a treatment effect?
- 2 estimands
  - Intent-to-treat effect (ITT)
  - Complier average causal effect ($CACE$)
- Choice of estimands depends, of course, on your research question and goal of causal inference

Intent-to-Treat Effect

$$\begin{aligned}\text{ITT} \equiv & E[Y_i(z=1)] - E[Y_i(z=0)] \\ = & E[Y_i(z=1, d(1))] - E[Y_i(z=0, d(0))]\end{aligned}$$

- ITT captures the average effect of being assigned to the treatment group regardless of the proportion of the treatment group actually treated
- Which causal inference method does this setup remind you of?

Complier Average Causal Effect

$$\text{CACE} \equiv \underbrace{E[(Y_i(d=1) - Y_i(d=0))}_{\text{average treatment effect}} \mid \underbrace{d_i(1) - d_i(0) = 1]}_{\text{among Compliers}}$$

Let

$$\pi_C = E[d_i(z=1) - d_i(z=0)]$$

be the proportion of compliers in the sample.

Then, the sample analog of the *CACE* estimand is

$$\text{CACE} = \frac{\text{ITT}}{\pi_C}$$

- Assumptions: Non-interference, excludability, and, under 2-sided noncompliance, monotonicity (no defiers, i.e. $d_i(1) \geq d_i(0)$)
- CACE also referred to as Local Average Treatment Effect (LATE) and, under one-sided noncompliance, Treatment on Treated (TOT)
- ATE among Compliers

Potential Outcomes

Obs	$Y_i(0)$	$Y_i(1)$	$D_i(0)$	$D_i(1)$	Type
1	4	6	0	1	Complier
2	2	8	0	0	Never-Taker
3	1	5	0	1	Complier
4	5	7	0	1	Complier
5	6	10	0	1	Complier
6	2	10	0	0	Never-Taker
7	6	9	0	1	Complier
8	2	5	0	1	Complier
9	5	9	0	0	Never-Taker

Compare ITT, ATE, and CACE

- ATE does not consider noncompliance:

$$\text{ATE} = \frac{2+6+4+2+4+8+3+3+4}{9} = 4$$

- ITT accounts for the fact that never-takers will not receive the treatment (always-takers will receive the treatment):

$$\text{ITT} = \frac{2+0+4+2+4+0+3+3+0}{9} = 2$$

- CACE is based on the subset of Compliers:

$$\text{CACE} = \frac{2+4+2+4+3+3}{6} = 3$$

Personal Canvass & Voting

- Gerber and Green New Haven study APSR 2000
- Randomly assign voters different GOVT tactics
  - Personal canvassing contact?
  - Mail?
  - Telephone?
  - Control?

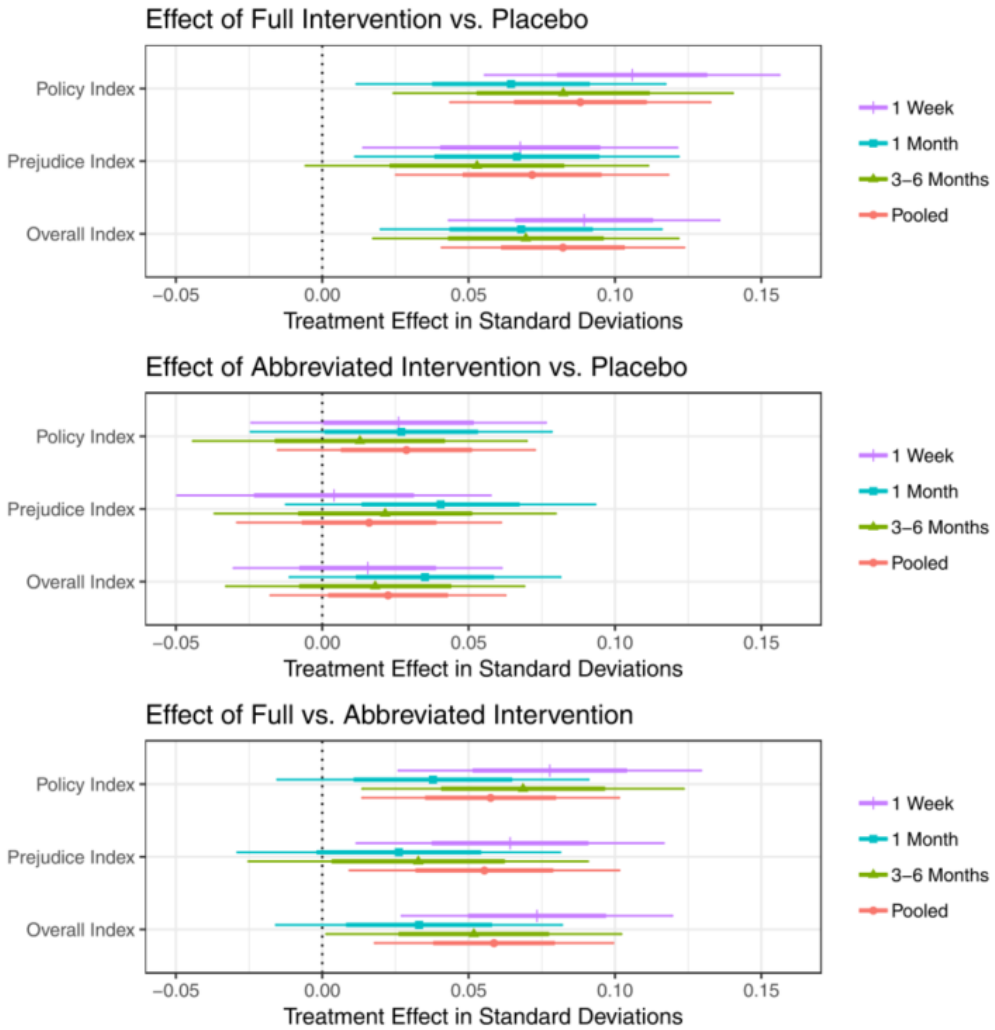
New Haven Voter Mobilization

Turnout Rate	Treatment Group	Control Group
Among those contacted	54.43 (395)	
Among those not contacted	36.48 (1050)	37.54 (5645)
Overall	41.38 (1445)	37.45 (5645)

- ITT = 41.38 - 37.54 = 3.84
- $\pi_C = 395/1445 = 0.273$
- CACE = ITT$/\pi_C = 3.84/0.273 = 14.1$

Kalla and Broockman (2020)

FIGURE 1. Experiment 1 Results: Intent-to-Treat Effects



Notes: Each panel shows the estimated intent-to-treat effects when comparing the two experimental conditions described in the panel title (e.g., the top panel compares the Full Intervention condition to the Placebo condition). Within each panel, we show treatment effects on the pre-specified primary outcome indices. Results are average treatment effects with 1 standard error (thick) and 95% confidence intervals (thin). To form each pooled index, we average each respondent's values for the corresponding index across all post-treatment survey waves. See Online Appendix Tables OA.9–11 for numerical point estimates and standard errors.

Broader takeaways

1. Carefully define the treatment itself
2. Carefully define treatment assignment and treatment receipt
3. Carefully define and try to identify compliant and non-compliant subgroups of subjects

Design implications

Bear in mind that

$$\text{SE}(\widehat{CACE}) \approx \frac{\text{SE}(\widehat{ITT})}{\pi_C}$$

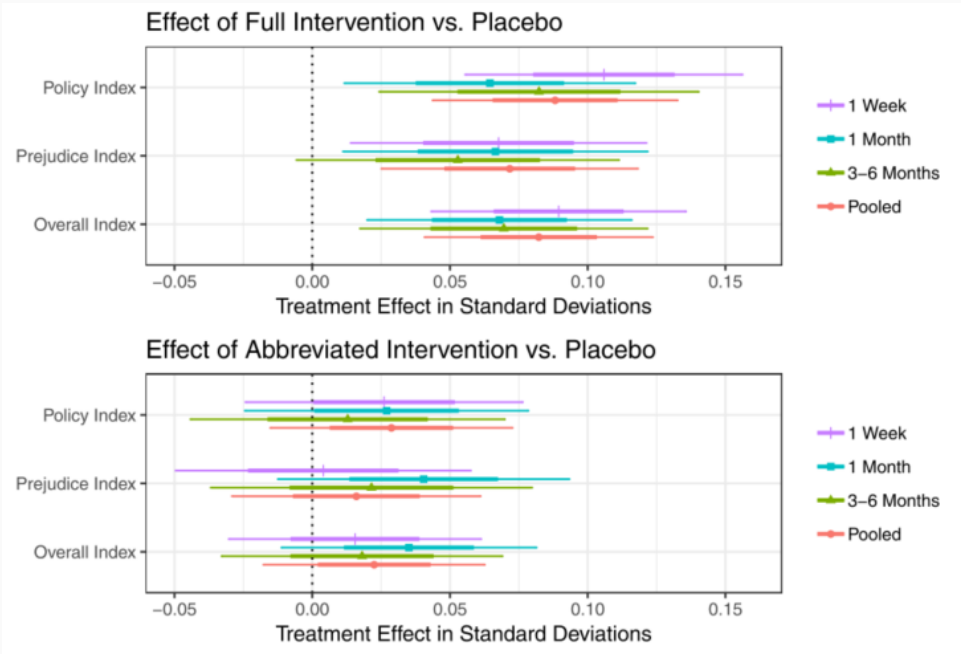
- Increase π_C; rule out defiers
- 1-sided noncompliance: Placebo design

Placebo Design

- Researchers attempt to contact individuals assigned to receive the treatment
- Those reached are then randomly allocated to two different groups
    - Treatment group
    - Placebo group receiving a "non-treatment"
- Kalla and Broockman (2020) canvassing experiment
    - Narratives (treatment)
    - Housing in Orange County (placebo)
- CACE estimated by comparing the outcomes for those in the treatment group to those in the placebo group
    - Random sample of Compliers whose untreated potential outcomes can be measured

Kalla and Broockman (2020)

Experiment 1		
Unauthorized immigrants		
Full Intervention	Abbreviated Intervention	
YES	NO	Positive effects ($d = 0.08$, $p < 0.001$)
NO	NO	Null effects ($d = 0.02$, $p = 0.27$), statistically distinguishable from Full Intervention ($d = 0.06$, $p < 0.01$)
YES	YES	$d = 0.12$
		$d = 0.03$ (Abbreviated vs. Placebo)



Placebo Design

- Logic is that placebo design screens out Never-Takers (since they are, in addition to compliers, part of the control group under 1-sided noncompliance)
- Compliers in the treatment group are compared directly to Compliers in the untreated group
- Reduces noise from Never-Takers in both treatment and control groups
- Moves us to a world of "full compliance"

Placebo Design

- Downside is that not all Compliers receive the treatment
- Resources are wasted on those receiving the placebo
- Opportunity to collaborate with someone studying an unrelated topic

Placebo Design

- The placebo and conventional design both allow estimation of the CACE
- Choice depends on the budget and compliance rate
- Under a fixed budget, the conventional design is preferable if compliance rate $> 50\%$ ($\pi_C > 1/2$)
- Canvassing studies often have a lower rate
- A pilot study may give a better idea of the expected compliance rate

Nickerson 2008

- Researchers attempt to contact individuals assigned to receive the treatment
- Those reached are then randomly allocated to two different groups
    - Treatment group
    - Placebo group receiving a "non-treatment"
- Nickerson (2008) canvassing experiment
    - GOTV (treatment)
    - Recycling (placebo)
- CACE estimated by comparing the outcomes for those in the treatment group to those in the placebo group
    - Random sample of Compliers whose untreated potential outcomes can be measured

Nickerson 2008

	Denver		Minneapolis		Pooled	
	Direct	Secondary	Direct	Secondary	Direct	Secondary
Percent Voting in	47.7%	42.4%	27.1%	23.6%		
GOTV Group	(3.0)	(2.9)	(3.1)	(3.0)		
Percent Voting in	39.1%	36.9%	16.2%	17.3%		
Recycling Group	(2.9)	(2.9)	(2.7)	(2.7)		
Estimated Treatment	**8.6%**	**5.5%**	**10.9%**	**6.4%**	**9.8%**	**6.0%**
Effect	**(4.2)**	**(4.1)**	**(4.1)**	**(4.1)**	**(2.9)**	**(2.9)**
P-Value	0.02	0.09	<0.01	0.06	<0.01	0.02

Note. Numbers in parentheses represent standard errors. P-values test the one-tailed hypothesis. Pooled estimates are weighted averages of results for both cities.

Nickerson 2008

- The placebo and conventional design both allow estimation of the CACE
- Choice depends on the budget and compliance rate
- Under a fixed budget, the conventional design is preferable if compliance rate $> 50\%$
- Canvassing studies often have a lower rate
- A pilot study may give a better idea of the expected compliance rate

Nickerson 2008