Blink, Buzz, Ring

A Randomized Experiment on Smartphone Notifications and Stress

Cendy Lin, Andrew Lam, Dominic Delmolino, Matthew Nelson
W241 Section 1
April 2017

Research Question

Will suppressing smartphone application notifications reduce the average individual's self-assessed stress level?

Multiple studies have shown that high stress levels negatively impact health
http://www.apa.org/helpcenter/stress.aspx

Few prior studies on phone stress, small populations (30 people): "The Do Not Disturb Challenge, 2016"
http://bit.ly/2pgvFj2

Very small studies (6 couples) report smartphone impacts on holiday sexual enjoyment:
https://campaignsoftheworld.com/outdoor/durex-donotdisturb-enjoy-great-holiday-sex/

Clearly an experiment is necessary...

Specific Hypothesis

Utilizing a smartphone's
"Do Not Disturb" feature for **application notifications**
(while retaining notifications for phone calls)
will reduce self-assessed stress levels while the feature is in use.

We expect reported stress levels **to go down**
for people in the experiment

Experimental Design

Subject identification and acquisition

Invitations to friends via Facebook, Slack and IRL (in real life)

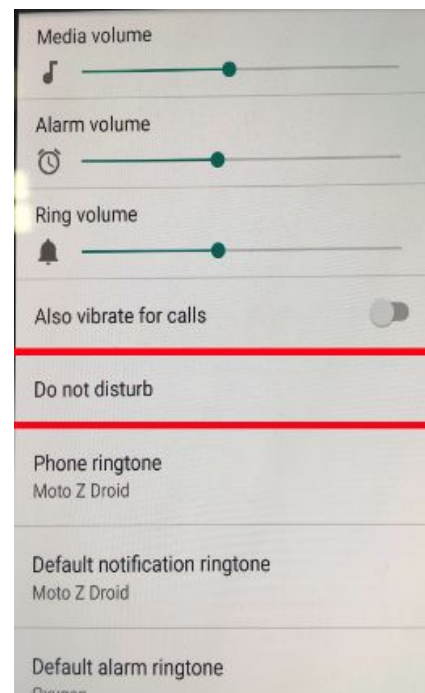Twitter ad campaigns calling for volunteers

Links to participant interest survey

Control and Treatment Interventions

Control users directed to *turn off* "Do Not Disturb" and to enable "Night Shift" (distraction)

Treatment users directed to *turn on* "Do Not Disturb" for *applications* and to allow phone calls.
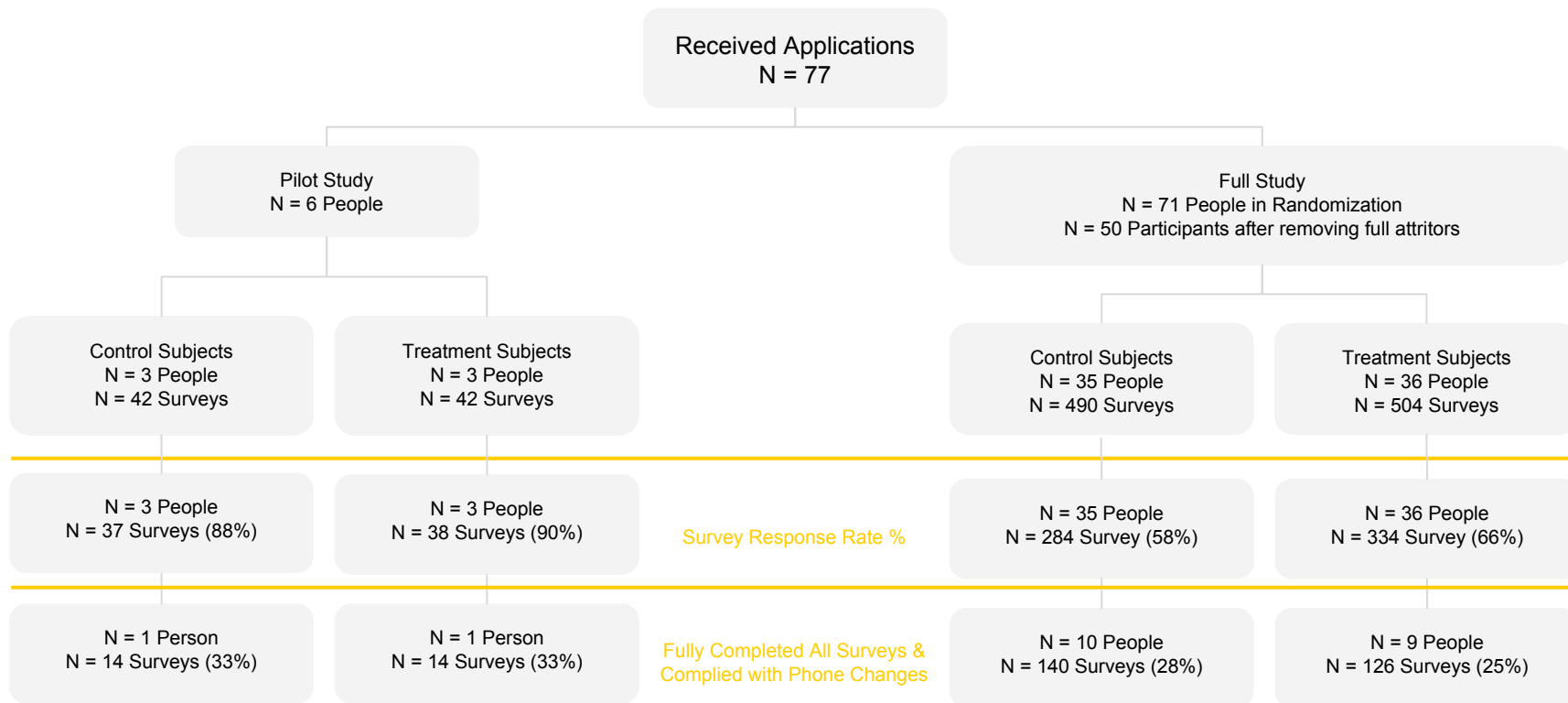
Instructions provided for iOS and Android

Measurement 'Units' and Randomization

- Daily Self-Assessment Surveys
  - Asked 'what was your stress level today?'
  - Asked 'distraction' questions relating to sleep, energy, etc.
  - Likert Scales from 1-5 (1 = "non-existent", 5 = "extremely high" levels of stress)
  - Directly asked about compliance
- Compliance levels dictated # of entries into the drawing for Amazon Gift Cards
- Sent out every evening, 24 hour response period requested
- 1 Week of Baseline Survey Responses
- 1 Week of Treatment Survey Responses
- Randomization Procedure
  - Blocked on gender, age, phone operating system, and number of phones for a total of 36 blocks
  - Subjects then randomly assigned to treatment and control within these blocks
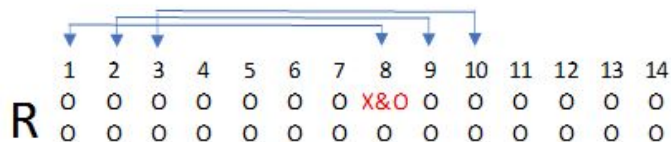 - Details described in flow diagram on next slide

Flow Diagram

Received Applications
N = 77

Pilot Study
N = 6 People

Full Study
N = 71 People in Randomization
N = 50 Participants after removing full attritors

Control Subjects
N = 3 People
N = 42 Surveys

Treatment Subjects
N = 3 People
N = 42 Surveys

Control Subjects
N = 35 People
N = 490 Surveys

Treatment Subjects
N = 36 People
N = 504 Surveys

N = 3 People
N = 37 Surveys (88%)

N = 3 People
N = 38 Surveys (90%)

Survey Response Rate %

N = 35 People
N = 284 Survey (58%)

N = 36 People
N = 334 Survey (66%)

N = 1 Person
N = 14 Surveys (33%)

N = 1 Person
N = 14 Surveys (33%)

Fully Completed All Surveys &
Complied with Phone Changes

N = 10 People
N = 140 Surveys (28%)

N = 9 People
N = 126 Surveys (25%)

Research Design

ROXO Notation

DiD Comparison for daily stress change from week-to-week

R
1	2	3	4	5	6	7	8	9	10	11	12	13	14
O	O	O	O	O	O	O	X&O	O	O	O	O	O	O
O	O	O	O	O	O	O	O	O	O	O	O	O	O

Example Schedule of Email Distribution

Week				1							2			
	Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
Assigned To:	4/3	4/4	4/5	4/6	4/7	4/8	4/9	4/10	4/11	4/12	4/13	4/14	4/15	4/16
Control	DS3	DS3	DS3	DS3	DS3	DS3	DS3	DS2	DS2	DS2	DS2	DS2	DS2	DS2
Control	DS3	DS3	DS3	DS3	DS3	DS3	DS3	DS2	DS2	DS2	DS2	DS2	DS2	DS2
Control	DS3	DS3	DS3	DS3	DS3	DS3	DS3	DS2	DS2	DS2	DS2	DS2	DS2	DS2
Treatment	DS3	DS3	DS3	DS3	DS3	DS3	DS3	DS1	DS1	DS1	DS1	DS1	DS1	DS1
Treatment	DS3	DS3	DS3	DS3	DS3	DS3	DS3	DS1	DS1	DS1	DS1	DS1	DS1	DS1
Treatment	DS3	DS3	DS3	DS3	DS3	DS3	DS3	DS1	DS1	DS1	DS1	DS1	DS1	DS1

DS3 = Daily Survey3 (Baseline Data Gathering)
DS1 = Daily Survey 1 (Treatment)
DS2 = Daily Survey 2 (Control)
Treatment Instructions
Control Instructions
Start of Experiment Instructions

7

Covariate Check

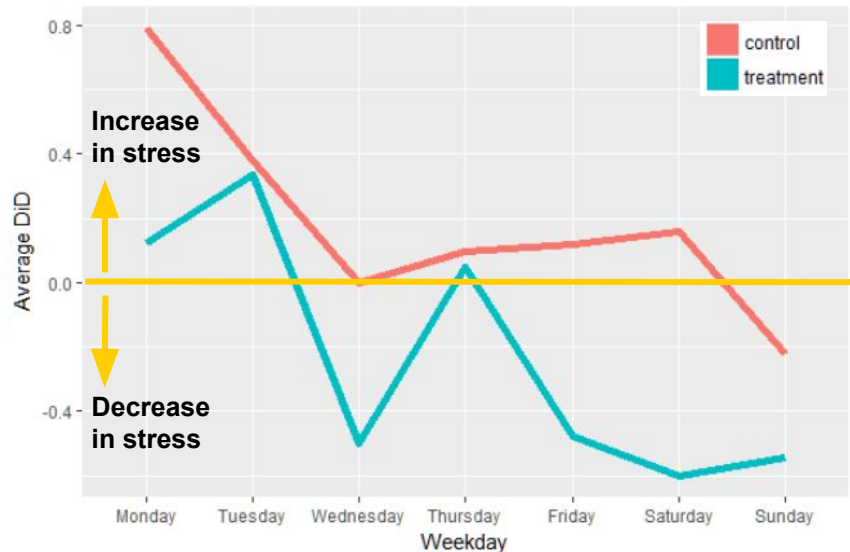| Covariate	Difference in Means		
	Control Mean 38 people	**Treatment Mean 39 people**	**P-value of Difference**
Male	0.47	0.51	0.735
Age (<44)	0.84	0.82	0.803
Apple Phone	0.79	0.85	0.526
Work+Personal phone (vs. personal phone only)	0.26	0.31	0.670
Phone Use (at least once per hour)	0.95	0.92	0.670
Relationship (Family/Friend vs. Other)	0.55	0.62	0.582

Change in Stress

- Difference in stress levels on the same day of the week (**Δstress**)
- **Example:** Treatment Monday (Day 8) - Baseline Monday (Day 1)

Distribution of Δstress by Group

Decrease in stress ← | → Increase in stress

N = 293
Missing = 246
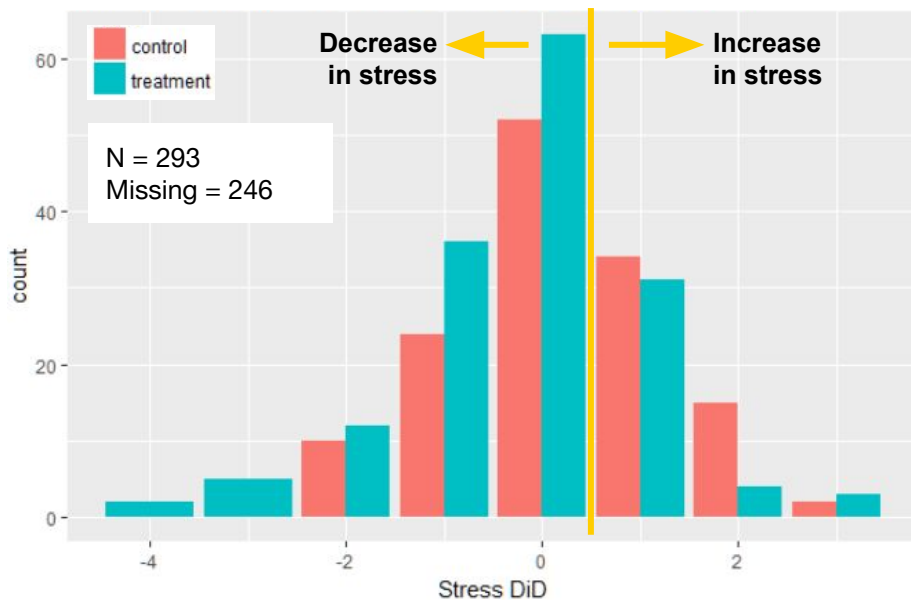
Average Δstress by Weekday and Group

Increase in stress

Decrease in stress

Initial Findings: Ignoring Attrition

	Dependent variable:	
	coefficient test Clustered SE (1)	OLS Fixed Effects (2)
treattreatment	-0.416*** (0.154)	-0.964 (0.619)
weekdayMonday	0.631*** (0.233)	0.677*** (0.259)
weekdaySaturday	-0.056 (0.186)	-0.088 (0.250)
weekdaySunday	-0.191 (0.219)	-0.180 (0.257)
weekdayThursday	0.257 (0.218)	0.240 (0.254)
weekdayTuesday	0.558** (0.282)	0.588** (0.251)
weekdayWednesday	-0.059 (0.295)	0.008 (0.255)
Constant	0.020 (0.188)	0.393 (0.453)
Fixed effects?	No	Yes
Observations		293
R2		0.298
Adjusted R2		0.121
Residual Std. Error		1.112 (df = 233)
F Statistic		1.679*** (df = 59; 233)

Note: *p<0.1; **p<0.05; ***p<0.01

$$\Delta Stress = b_0 + b_1 treatment + b_i weekday_i$$

- **N = 293 of 539 (Δstress observations)**

- **Non-Compliance:** 33% of *observed* responses in treatment group, $\alpha = 0.67$

- **CACE**
  - **Clustered SE:** -0.416/0.67 = **-0.62**
  - **Individual fixed effects:** -0.964/0.67 = **-1.43**

- **Attrition:**
  - Missing 46% of Δstress observations across control and treatment groups
  - Cannot assume MIPO from experiment design

A Look at Attrition

Attrited DiD Distribution by Treatment Group
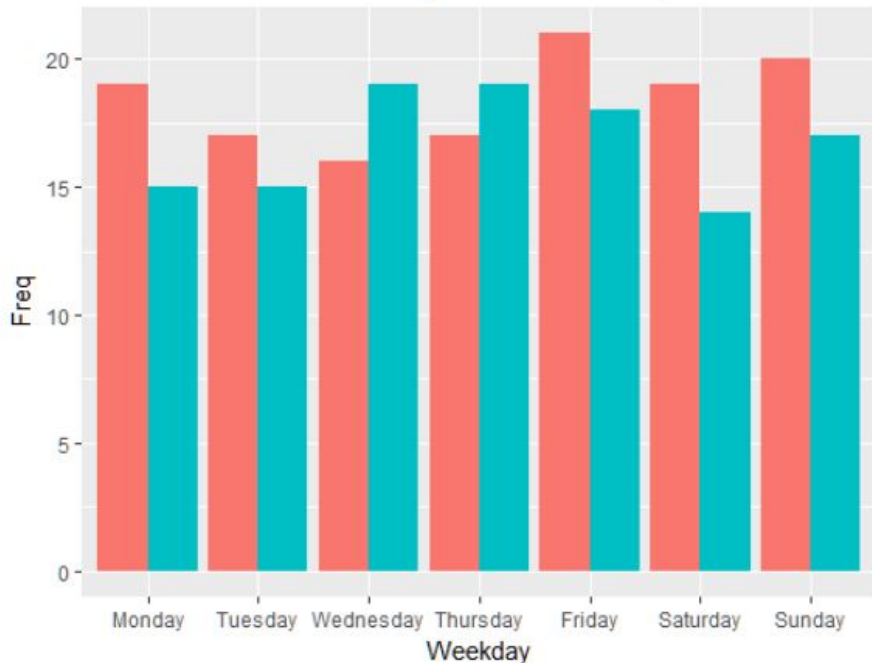
N= 539 (DiD or Δstress observations)

- **46% missing** Δstress observations overall
- **48% missing** Δstress observations in control
- **44% missing** Δstress observations in treatment
- Check for randomness on observed responses against prognostic variables **failed** (obtained a significant result on F-test):

F-Test Check for Randomness
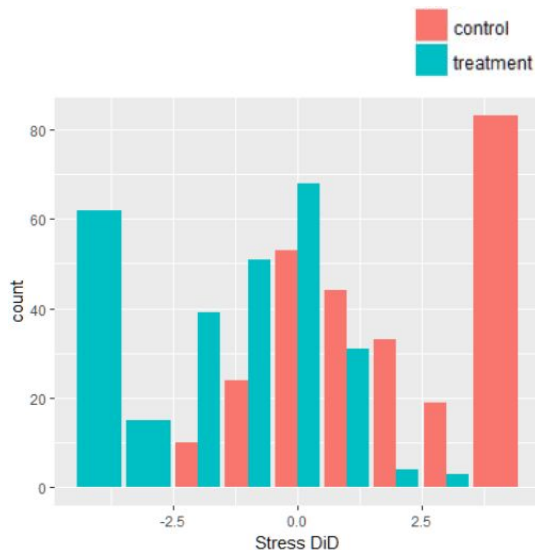
```
Model 1: treat_code ~ 1
Model 2: treat_code ~ 1 + male + age + rel + apple + pers + use
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       292     404.95
2       286     381.10  6    23.85  0.0005565 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Attrition: Banded Range - Max Effect

Max Effect:
- Control Δstress: +4
- Treatment Δstress: -4

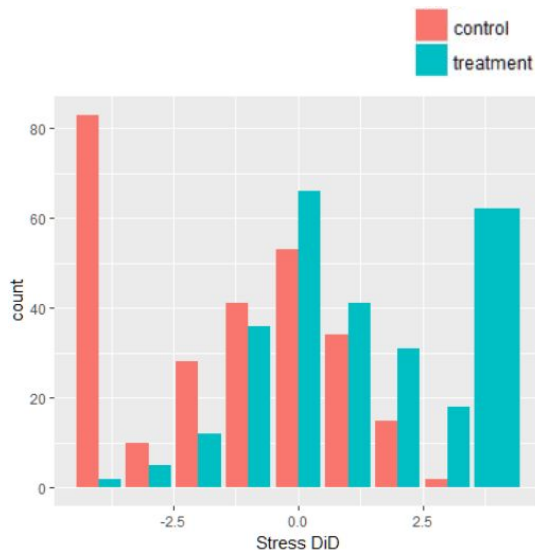|	Dependent variable:	
	coefficient test Clustered SE (1)	max_did OLS Fixed Effects (2)
treattreatment	-3.080*** (0.347)	-4.571*** (0.626)
weekdayMonday	0.273 (0.219)	0.273 (0.189)
weekdaySaturday	-0.026 (0.146)	-0.026 (0.189)
weekdaySunday	-0.221 (0.193)	-0.221 (0.189)
weekdayThursday	0.026 (0.176)	0.026 (0.189)
weekdayTuesday	0.104 (0.202)	0.104 (0.189)
weekdayWednesday	-0.195 (0.196)	-0.195 (0.189)
Constant	1.716*** (0.287)	4.006*** (0.460)
Individual fixed effects?	No	Yes
Observations		539
R2		0.800
Adjusted R2		0.764
Residual Std. Error		1.172 (df = 456)
F Statistic		22.295*** (df = 82; 456)

Note: *p<0.1; **p<0.05; ***p<0.01

CACE

- **Clustered SE:**
  -3.08/0.67
  = **-4.59**

- **Individual fixed effects:**
  -4.57/0.67
  = **-6.82**

Attrition: Banded Range - Min Effect

Min Effect:
- Control Δstress: -4
- Treatment Δstress: +4


```
===========================================
                   Dependent variable:
                  -------------------------
                  coefficient      OLS
                     test
                  Clustered SE  Fixed Effects
                      (1)          (2)
-------------------------------------------
treattreatment     2.642***      4.000***
                   (0.385)       (0.661)

weekdayMonday       0.169         0.169
                   (0.211)       (0.199)

weekdaySaturday    -0.234        -0.234
                   (0.174)       (0.199)

weekdaySunday      -0.221        -0.221
                   (0.207)       (0.199)

weekdayThursday     0.234         0.234
                   (0.192)       (0.199)

weekdayTuesday      0.260         0.260
                   (0.235)       (0.199)

weekdayWednesday    0.013         0.013
                   (0.224)       (0.199)

Constant           -1.494***     -4.032***
                   (0.296)       (0.485)
-------------------------------------------
Individual fixed effects?  No       Yes
Observations                        539
R2                                  0.779
Adjusted R2                         0.739
Residual Std. Error      1.237 (df = 456)
F Statistic        19.579*** (df = 82; 456)
===========================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

CACE

- **Clustered SE:**
  2.64/0.67
  = **3.94**

- **Individual fixed effects:**
  4/0.67
  = **5.97**

Issues with the Experiment

1. **Self-response Surveys**
   a. **Attrition**
      i. 46% attrited DiD responses
      ii. Used upper, lower bounds to find banded range
   b. **Non-compliance**
      i. 33% non-compliant non-attrited observations in treatment
      ii. Do we trust participant responses?
   c. **Survey Design**
      i. Does not indicate which day the survey was for - made assumptions here
      ii. Collect time zone and time of day of response (control for AM/PM response)

2. **Generalizability**
   a. 71% of subjects were friends/family, >90% acquired via social media, can't generalize to broader populations
   b. Results only applicable to the types of people willing to make changes to their smartphone notifications and opt into this experiment

Questions and Concerns

- **Panel Data**
  - Use fixed effects or clustered standard errors on individuals?

- **Attrition**
  - How to deal with attrition if people only responded sometimes?
  - Could potentially go back to collect data but limited on time

- **Non-compliance**
  - How to determine non-compliers if people complied sometimes?
  - Did people actually follow instructions? Can't tell because no way of monitoring except questions
  - One-sided non-compliance only

- **Non-persistence**
  - How does analysis change if we don't assume non-persistence of treatment?

Lessons Learned

<u>What did you learn?</u>

- Suppressing notifications may or may not reduce stress -- follow up study needed!
  - Promising but too much attrition
- Key elements of experiment design
- Strength of blocking for random assignment
- Clear instructions, interventions for control (placebo) and treatment
- Power of difference-in-differences as a control measurement
- Difficulty in securing survey participation on a regular basis
- Difficulty in acquiring subjects
- Potential of Twitter for random subject acquisition (subject to population bias of who's on Twitter)
- Value of a pilot study

<u>What might you have done differently given infinite time/resources?</u>

- Get more participants from a broader pool of people
- Use biometric markers of stress vs. self-survey (e.g. heart rate, cortisol levels, etc.)
- Provide better incentives for people to respond or better yet, find a way to automatically capture data
- Monitor phones for compliance (e.g. through an app)
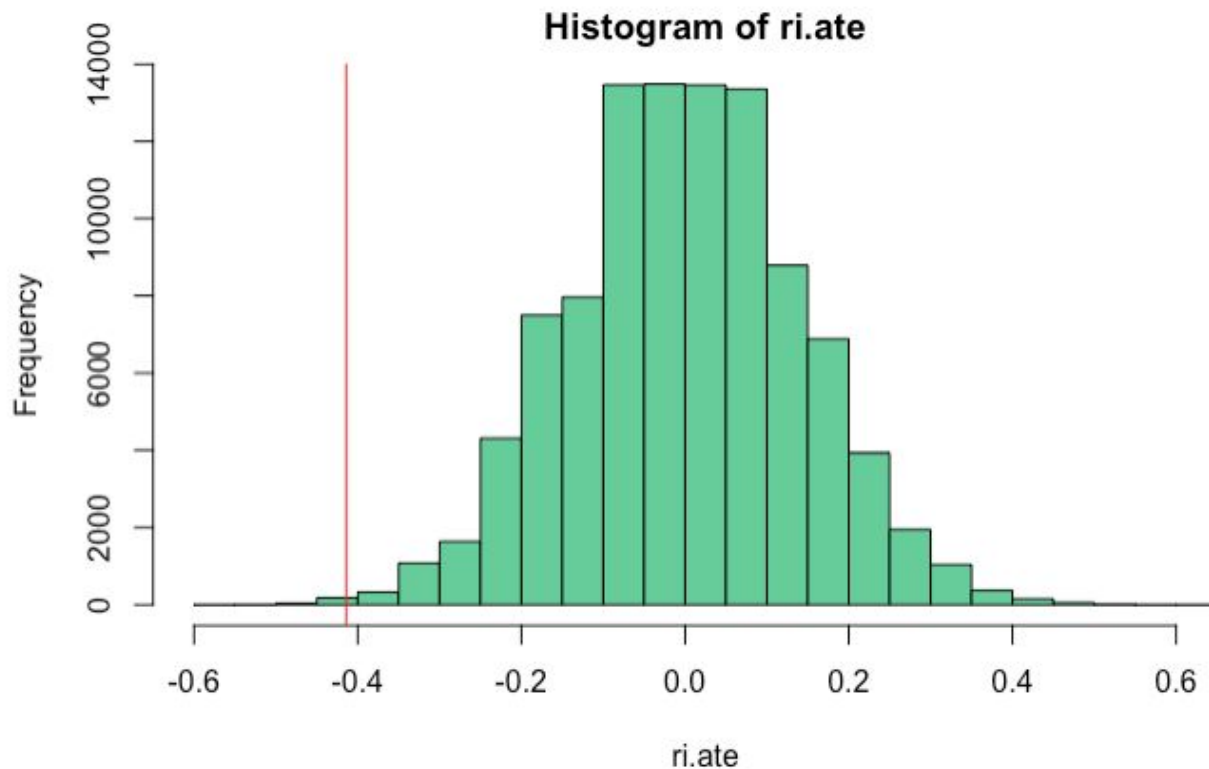- Longer baseline and treatment / control periods
- Long-term effects

Questions?

Appendix

Randomization Inference

Histogram of ri.ate

Presentation Guidelines

- During one of the final classes, we will ask you to present your findings to your peers for feedback that might help you improve your final paper. Please don't spend time making the presentation pretty; this will not get an explicit grade.
- Present experiment design and preliminary results to classmates, take feedback from peers, build forward toward final analysis, generate ideas for further research
- Format:
  - 15-20 minutes presentation
  - 5-10 minutes for questions and feedback
  - What is the experimental question
  - why is it interesting/important (cite prior literature if relevant)
  - how did you seek to test the question
  - what were the main results?
  - what were some limitations of your experiment? what did you learn? what might you have done differently given infinite time/resources?
  - presentations will take place the last two weeks of class