How to Randomize
Course overview

1. Why Evaluate
2. Theory of Change & Measurement
3. Why & When to Randomize
4. How to Randomize
5. Sample Size & Power
6. Randomized Evaluation from Start to Finish
7. Threats & Analysis
8. Ethical Considerations
9. Generalizing & Applying Evidence
Learning Objectives

• Develop a deeper understanding of how randomization works using real-world examples
• Determine how to select the appropriate level of randomization
• Review common program specifications and how these inform randomization design choices
Lecture Overview

• What is Randomization?
• Randomization Process & Design
• Balance & Stratification
• The Unit of Randomization
• Designing Randomized Evaluations for Different Program Specifications
Lecture Overview

• What is Randomization?
• Randomization Process & Design
• Balance & Stratification
• The Unit of Randomization
• Designing Randomized Evaluations for Different Program Specifications
Random Sampling and Random Assignment

Defining terms

**Random sampling**: Selecting units from a population of interest in a randomized manner to create a sample that is representative of the population.

**External validity**: The acceptability of results of an evaluation in contexts other than those in which the evaluation was conducted.

**Random assignment**: Taking a pool of eligible units and then allocating those units to treatment and control groups by means of a random process.

**Internal validity**: The acceptability of the results of an evaluation in terms of causal impact of the intervention.
Random Sampling

Randomly sample from area of interest
Random Assignment

Randomly assign to treatment and control

Google Maps
Random Assignment

Monthly income, per capita ($)

<table>
<thead>
<tr>
<th>Population</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>6,404</td>
<td>6,458</td>
<td>6,367</td>
</tr>
</tbody>
</table>

Google Maps
A real-world example

Randomization of the program across zip codes in 13 states

Effects of large-scale social media campaign on holiday travel and Covid-19

- **Context & intervention:** Before peak holiday travel in 2020, we shared brief videos over social media of medical providers encouraging people to stay home for the holidays, with the goal of reducing the spread of Covid-19

- **Sample selection**
  - Unit of randomization: US zip codes
  - Sample frame: All US states reporting weekly Covid-19 case count data at the zip code level

Randomization before Thanksgiving and before Christmas

Lecture Overview

• What is Randomization?
• Randomization Process & Design
• Balance & Stratification
• The Unit of Randomization
• Designing Randomized Evaluations for Different Program Specifications
Randomization can be appropriate when a program is oversubscribed and resource constraints prevent everyone who is eligible from receiving the treatment.

A “lottery” can be easy to understand and implement:
- Randomly assign treatment and control from sample of those who are eligible (example: some housing authorities run lotteries for Housing Choice Vouchers).
“Rolling Randomization” Process

Randomize each participant upon entry into study/at the point of service

- Set the probability of assignment to treatment group to a fixed percentage (e.g., 50%, 75%)

<table>
<thead>
<tr>
<th>ID</th>
<th>Coin</th>
<th>Treatment/Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Heads</td>
<td>T</td>
</tr>
<tr>
<td>2</td>
<td>Heads</td>
<td>T</td>
</tr>
<tr>
<td>3</td>
<td>Tails</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>Heads</td>
<td>T</td>
</tr>
<tr>
<td>5</td>
<td>Tails</td>
<td>C</td>
</tr>
<tr>
<td>6</td>
<td>Heads</td>
<td>T</td>
</tr>
<tr>
<td>7</td>
<td>Tails</td>
<td>C</td>
</tr>
<tr>
<td>8</td>
<td>Tails</td>
<td>C</td>
</tr>
<tr>
<td>9</td>
<td>Heads</td>
<td>T</td>
</tr>
<tr>
<td>10</td>
<td>Heads</td>
<td>T</td>
</tr>
</tbody>
</table>

Count: T: 6, C: 4
Varying Intensity of Treatment

This type of research design can help answer the question: “What is the effect of varying the intensity of the treatment?”

At the individual level (i.e., varying the dosage each individual gets)

At the group/cluster level (i.e., varying the proportion of treated people within an area)
Multiple treatments

How do different treatments compare?

<table>
<thead>
<tr>
<th>Covid-19 info via video</th>
<th>Covid-19 information via text message</th>
<th>No messaging</th>
</tr>
</thead>
</table>

How do different treatments interact?

<table>
<thead>
<tr>
<th>Video</th>
<th>Text message</th>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Segment 3</th>
<th>Segment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Text message + video</td>
<td>Video</td>
<td></td>
<td>Control</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>Group 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Text message</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
<th>Group 2</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Video</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
<th>Group 3</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Text message</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Impact of Tailored Public Health Messages for Black and Latinx Communities on Knowledge and Behavior

Context: Black and Latinx communities disproportionately affected by Covid-19 due to systemic racism and public health messaging that may not be framed to address these communities’ concerns

Intervention: Three videos about generic health topics or Covid-19
- Delivered by a physician of the same or different race/ethnicity of the viewer
- With tailored acknowledgement of injustice or with standard statement

Effects of Tailored Messaging for Black and Latinx Communities on Covid-19 Knowledge and Behaviors

<table>
<thead>
<tr>
<th>Tailored acknowledgement</th>
<th>Racial concordance + tailored acknowledgement</th>
<th>Racial concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Group 1</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Racial concordance</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Group 3</td>
<td>Group 4</td>
</tr>
<tr>
<td></td>
<td>Racial concordance</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lecture Overview

• What is Randomization?
• Randomization Process & Design
• Balance & Stratification
• The Unit of Randomization
• Designing Randomized Evaluations for Different Program Specifications
Balance & Stratification

**Balance** = the treatment and control groups are comparable on certain key characteristics

- Can check using **balance tests**. If you find that some variables are unbalanced, consider the number of imbalances and their magnitude.

**Stratification** = dividing the sample into **subgroups** (also known as **strata**) that share certain characteristics and then randomizing within each subsample
Stratification

Steps:
1. Form groups of similar units
2. Randomize within the group

Example:
1. Divide sample into age groups
2. Randomize individuals separately within each education level
Stratification

Steps:
1. Form groups of similar units
2. Randomize within the group

Example:
1. Divide sample into age groups
2. Randomize individuals separately within each education level
Stratification

Steps:
1. Form groups of similar units
2. Randomize within the group

Example:
1. Divide sample into age groups
2. Randomize individuals separately within each education level
Stratification

Randomization sets up an expectation of balance. Stratification ensures it.

But only for this one observable characteristic!
Why stratify?

1. To **achieve balance**
2. To **increase statistical power** (we will return to this tomorrow)
3. To analyze treatment effect by group (also known as **subgroup analysis** or heterogeneous treatment effects)
   - For example, differential effects by race or ethnicity, rural/urban settings, income level, etc.

During random sampling, in some cases, we may want to use stratification to ensure certain groups are overrepresented (**disproportionate stratification**)
Impact of Tailored Public Health Messages for Black and Latinx Communities on Knowledge and Behavior

**Intervention**: Three videos about generic health topics or Covid-19, varied by racial concordance and messaging

**Sample and stratification**:  
– **Eligible sample**: 18 years or older, self-identifying as Black or Latinx, oversampled with a secondary education or less  
– **Stratified** by sex, age (45 years and older), and location

**Results**: Physician messaging campaign increased Covid-19 knowledge and self-reported protective behaviors. Racial concordance or tailored messaging did not further increase effectiveness of messaging.

## Baseline characteristics/balance check

### Appendix Table. Summary of Participant Characteristics*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study Participants</th>
<th>Control</th>
<th>Any Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>African American</td>
<td>Latinx</td>
</tr>
<tr>
<td>Observations, n</td>
<td>14144</td>
<td>8767</td>
<td>5377</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8112</td>
<td>5277</td>
<td>2835</td>
</tr>
<tr>
<td>Male</td>
<td>6032</td>
<td>3490</td>
<td>2542</td>
</tr>
<tr>
<td>Mean age (SD), y</td>
<td>38.93 (15.3)</td>
<td>39.63 (15.7)</td>
<td>37.78 (14.4)</td>
</tr>
<tr>
<td>Stratum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female, age ≥45 y, coastal</td>
<td>1318</td>
<td>829</td>
<td>439</td>
</tr>
<tr>
<td>Female, age ≥45 y, central</td>
<td>2013</td>
<td>1245</td>
<td>765</td>
</tr>
<tr>
<td>Female, age ≤44 y, coastal</td>
<td>2010</td>
<td>1245</td>
<td>765</td>
</tr>
<tr>
<td>Female, age ≤44 y, central</td>
<td>2771</td>
<td>1773</td>
<td>998</td>
</tr>
<tr>
<td>Male, age ≥45 y, coastal</td>
<td>820</td>
<td>521</td>
<td>299</td>
</tr>
<tr>
<td>Male, age ≥45 y, central</td>
<td>996</td>
<td>641</td>
<td>355</td>
</tr>
<tr>
<td>Male, age ≤44 y, coastal</td>
<td>1939</td>
<td>1013</td>
<td>926</td>
</tr>
<tr>
<td>Male, age ≤44 y, central</td>
<td>2277</td>
<td>1315</td>
<td>962</td>
</tr>
<tr>
<td>Mean household size (SD), n</td>
<td>3.22</td>
<td>3.04</td>
<td>3.5</td>
</tr>
<tr>
<td>Mean education index (SD)†</td>
<td>4.56</td>
<td>4.69</td>
<td>4.35</td>
</tr>
</tbody>
</table>

* Appendix Table. Summary of Participant Characteristics*

What subgroup analyses are you interested in?
What variables might you want to stratify by?
Lecture Overview

• What is Randomization?
• Randomization Process & Design
• Balance & Stratification
• The Unit of Randomization
• Designing Randomized Evaluations for Different Program Specifications
Units of Observation and the Level of Randomization

At which level should a study randomize?

• Randomizing at the individual level
  - e.g., people, patients, or students

• Randomizing at the group/cluster level
  - e.g., villages, clinics, or schools
  - *Outcomes can still be measured at the individual level*
Unit of Randomization: Student?
Unit of Randomization: Student?
We call groups of units “clusters” Randomization at the group level: “cluster randomized trial”
Unit of Randomization: Classroom

We call groups of units “clusters” Randomization at the group level: “cluster randomized trial”
Unit of Randomization: School

We call groups of units “clusters”.
Randomization at the group level: “cluster randomized trial”
Unit of Randomization: School

We call groups of units “clusters” Randomization at the group level: “cluster randomized trial”
Level of Randomization: Considerations

• What is the unit of measurement?
  - What are the outcomes we care about, and at what level are we able to measure them?

• What unit does the program target for treatment?
  - At what level is the intervention administered? (e.g., classrooms for educational programs, factory lines for workforce projects)

• Which level gives us the greatest probability of detecting an effect? (This involves statistical power, which is the focus of tomorrow’s first session)

• Reality check: Which level is feasible ethically, financially, politically, and logistically?
Returning to Covid-19 social media campaign example

What are some considerations and constraints for the level of randomization in your contexts?
Level of Randomization: Considerations

Choose a level of randomization to minimize **noncompliance** and to measure or contain **spillovers** (these are two different concepts)

**Noncompliance:** When participants do not follow (“comply with”) their treatment assignment. For example, if participants assigned to the control group join the treatment group, or vice versa.

**Spillover:** When the treatment indirectly affects those who have not been treated. Spillover effects can be positive or negative.
Measuring Spillovers in Covid-19 Social Media Campaign

- Investigating spillover in high-intensity zip codes to untreated zip codes vs. spillover in low-intensity zip codes to untreated zip codes

Measuring Spillovers in Covid-19 Social Media Campaign

- Investigating spillover in high-intensity zip codes to untreated zip codes vs. spillover in low-intensity zip codes to untreated zip codes

Lecture Overview

- What is Randomization?
- Randomization Process & Design
- Balance & Stratification
- The Unit of Randomization
- Designing Randomized Evaluations for Different Program Specifications
Program Specifications

Evaluating programs with certain specifications requires adjustments. For example, adjustments might be required if:

• Eligibility is determined by a cutoff (e.g., income, credit score)
• There are enough resources to extend the program to everyone in the study area
• The intervention is a current entitlement program, and participation cannot and/or should not be withheld from anyone who is eligible
Program Specification:

Program has Strict Eligibility Criteria
Program with an Eligibility Cut-off

People

Eligible

Cut-off

Ineligible

Income
RCT Design: Expand Eligibility and Randomize among the Newly Eligible
RCT Design: Expand Eligibility and Randomize among the Newly Eligible
Program Specification:

Resources Exist to Extend the Program to Everyone in the Study Area
If it is not possible to randomize access to a program because you cannot withhold treatment from the control group, you might be able to conduct a randomized controlled trial by:

A. Randomizing timing of access to program
B. Randomizing encouragement to take up the program
C. Either A or B
D. Not sure
If it is not possible to randomize access to a program because you cannot withhold treatment from the control group, you might be able to conduct a randomized controlled trial by:

A. Randomizing timing of access to program
B. Randomizing encouragement to take up the program
C. Either A or B – depending on the context, either of these options might be appropriate.
D. Not sure
RCT Design: Phase-in
Phase 0: No One Treated Yet
All Control
Phase 1: 25% Treated
75% Control
Phase 2: Half Treated

Half Control
Phase 3: 75% Treated
25% Control
Phase 4: All Treated
No Control (Experiment Over)
Program Specification:

Entitlement Programs
Entitlement Programs: Cannot Mandate nor Deny Intervention
RCT Design: Encouragement

Sign up!

Treatment Group

Control Group
RCT Design: Encouragement

Sign up!
RCT Design: Encouragement

Treatment Group

Control Group

75% take-up

25% take-up

J-PAL | HOW TO RANDOMIZE
RCT Design: Encouragement

Treatment Group
75% take-up

Control Group
25% take-up

Compare Entire Treatment Group to Entire Control Group

Evaluation: https://www.povertyactionlab.org/evaluation/snap-take-evaluation
How might expanding eligibility, phase-in design, and encouragement design be of use in your contexts?
A Very Quick Recap

• What is Randomization?
• Randomization Process & Design
• Balance & Stratification
• The Unit of Randomization
• Designing Randomized Evaluations for Different Program Specifications
References


Appendix
How to randomize

Population of interest:
[Type answers here.]

Method of sampling and/or enrolling participants (or groups of participants):
[Type answers here.]
Number of units that will be included in the evaluation (if known):
[Type answers here.]

Randomization process
(e.g. randomize at the point of service, or randomize from a complete list of participants)
[Type answers here.]

Treatment group 1
Intervention that treatment group 1 receives:
[Type answers here.]
Number of units in treatment group 1 (if known):
[Type answers here.]

Treatment group 2
(if more than one treatment arm)
Intervention that treatment group 2 receives:
[Type answers here.]
Number of units in treatment group 2 (if known):
[Type answers here.]

Control group
Services ("business as usual") that the control group receives, if any:
[Type answers here.]
Number of units in the control group (if known):
[Type answers here.]
Which variables to use for stratification

• Variables that are **discrete**
  – Practical constraint (it is challenging to create “buckets” out of continuous variables)

• Variables you believe might be **correlated with the outcome of interest** or possible shocks
  – For example, female headed households have different access to inputs

• Variables for which you are interested in **heterogeneous (differential) treatment effects**
Unbalanced groups

What do you do if you conduct a balance test and find that some variables are unbalanced?

• Consider the number of variables that are unbalanced (remember that if you are testing at a 5% significance level, then you would expect to see differences between the groups roughly 5% of the time)

• What is the magnitude of difference? (a large magnitude of difference is more concerning than a small one)

• Which variables are unbalanced? (Imbalances can be especially problematic if the variables are correlated with take-up of treatment, attrition, or the outcome variables)
## Recap: Randomization designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Most useful when:</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic lottery</td>
<td>Program is oversubscribed/not enough resources to deliver to all eligible</td>
<td>Familiar and well understood</td>
<td>Control group may be less interested in answering questions/being involved, which could create problems for the analysis (differential attrition)</td>
</tr>
<tr>
<td></td>
<td>Resources are constant for the evaluation period</td>
<td>Easy to implement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Often seen as fair</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transparent process</td>
<td></td>
</tr>
</tbody>
</table>
Recap: Randomization designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Most useful when:</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized phase-in/roll out</td>
<td>Resources to administer program expands over time</td>
<td>Easy to understand</td>
<td>Anticipation of receiving the program may impact short-run behavior for the control group</td>
</tr>
<tr>
<td></td>
<td>Everyone must receive treatment eventually</td>
<td>Constraint is easy to explain</td>
<td>Only possible for outcomes that can be measured before everyone is treated</td>
</tr>
<tr>
<td></td>
<td>Logistically difficult to roll out to everyone at one time</td>
<td>Control group likely to comply because they expect to benefit later</td>
<td>Difficult to measure long-term impact</td>
</tr>
</tbody>
</table>
Recap: Randomization designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Most useful when:</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encouragement designs</td>
<td>Program must be open to everyone in target population</td>
<td>Allows evaluation of programs that cannot exclude anyone</td>
<td>Measures impact of those who respond to the encouragement (not the general population)</td>
</tr>
<tr>
<td></td>
<td>Take-up is low in the relevant population</td>
<td>Can randomize at individual level even if the program is administered at a group level</td>
<td>Need large enough inducement to improve take-up</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Encouragement itself may have direct effect on outcomes</td>
</tr>
</tbody>
</table>
## Recap: Randomization designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Most useful when:</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expand eligibility and randomize near an eligibility cutoff</td>
<td>There are enough funds to expand the number of people served by an intervention. The intervention is essential for certain people who are already being served</td>
<td>Everyone previously eligible for a program remains eligible.</td>
<td>Does not measure the intervention's effect on people who were already eligible. Only can measure effects for people who are near the eligibility cutoff.</td>
</tr>
</tbody>
</table>

J-PAL | HOW TO RANDOMIZE
Adaptive experiments

• A more flexible form of assignment

• Use information on treatment outcomes from earlier waves of the study to place more participants over time in the treatment conditions that are working the best
  - Example: algorithm placed participants in job assistance interventions that were showing to be most effective

• Enables quicker response to research findings

Resource:
• Adaptive experiments for policy research (VoxDev)

J-PAL studies:
• Adaptive Trial to Identify Maximally Effective COVID Information Campaign
• An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan
Additional J-PAL resources

- Randomization (research resource)
  - Real-world challenges to randomization and their solutions
  - Sample size and calculating statistical power (research resource)
  - Rules of thumb for statistical power
Reuse and citation

To reference this lecture, please cite as:


This lecture is made available under a Creative Commons Attribution 4.0 License (international): https://creativecommons.org/licenses/by/4.0/