Randomization and Stratification

Anja Sautmann

June 10, 2025

Agenda

1. Introduction: basics and definitions
2. Random assignment
3. Adherence to randomization

1 Introduction

Sampling and randomization
for unbiased effect estimates

Sampling and Randomization

- Sampling describes the process of (randomly) selecting units that are actually included in the study (the study sample)
- **Randomization** describes the process of assigning units to the different study arms (e.g. treatment and control)

The goal is to create treatment arms that are representative of the underlying population to be able to estimate unbiased treatment effects.

→ random *sampling* from the population (external validity)

→ *randomized assignment* to arms (internal validity)

… plus measures to increase precision, e.g. **stratification**.

In practice, samples are often not representative of the population of interest (TBC).

Unit of Analysis

- The **unit of analysis** is the “unit of interest”: unit for which treatment effects are expected

- Example: intervention at health clinics improving quality of care through training
  - Unit of analysis 1: providers; measure quality of care delivered
  - Unit of analysis 2: patients; measure satisfaction with care and health outcomes

- Typical units of analysis are households, workers, firms... but also: production line in a factory, mayors attending a conference, clinics, villages

- *Notes for later:*
  - *Multiple UoA* possible
  - UoA may not be the respondent/data source (e.g. principal vs. school)
  - Need precision: for whom or what were outcomes measured?

Unit of Randomization

- The **unit of randomization** is the unit which is assigned randomly to study arm.
- Often identical to the unit of analysis – but not always

- In more complex designs, the unit of randomization is a *cluster*, or connected group, of units of analysis.
  - Example: UoR is the village, UoA is the household

- Reason may be logistical or ethical - impossible to assign treatment at the individual level
- More often: reason is spillover effects across UoA

The unit of analysis and unit of randomization effect…

Outcome variable unit of measurement	Outcomes (Day 1 & Day 2)
Precision	Treatment effect estimates and precision (Day 3)

2

Random Assignment and Stratification

How are units of randomization assigned to arms?

Randomization

Goals when randomizing treatment assignment:

- Create study arms that are not systematically different, minimize random variation
  → **if treatment has no effect, difference between study arms (in expectation) a *precisely estimated* zero**

- Make sure SUTVA is satisfied: any unit not affected by treatment received by other units (spill-overs, externalities, experimental effects, etc.)
  → **measure the outcome effects that would occur outside of an experiment**

Cluster randomization

- **Cluster randomization:** groups of the units of analysis assigned to study arms
  - Groups that are close geographically or institutionally (villages, classrooms) → potential for spillovers between members

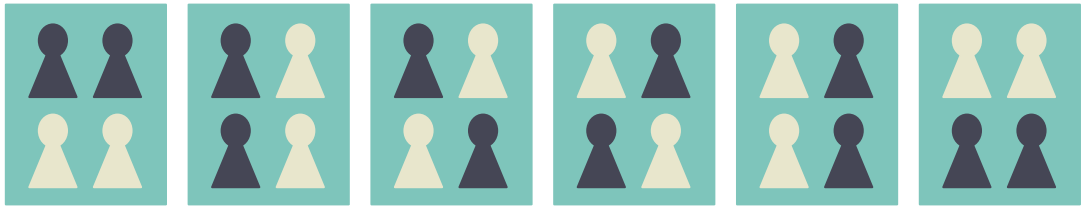
- Goal: avoid spillovers/SUTVA violations
  - Externalities/spillovers (example information interventions)
  - General equilibrium effects (e.g., intervention changes market prices)
  - Non-compliance to treatment assignment (contamination)
  - Experimental effects (e.g. John Henry effect)

- Notes: adjust standard errors for cluster randomization

Creating equal-sized treatment arms

- **Block randomization** or **permuted block randomization**
  - General process:
    - Create blocks (or lists) of UoR that are a multiple of the number of study arms; e.g. pairs or sets of four for two study arms
    - Create all possible permutations of assigning the block units to arms
    - Randomly select a permutation for each block.

- Equivalent (static setting): **randomly draw a fixed share of the sample**
  - Ex: randomly order all subjects, assign first half to treatment and second half to control (= 1 block, 1 permutation).
  - “Drawing treatment arms without replacement”

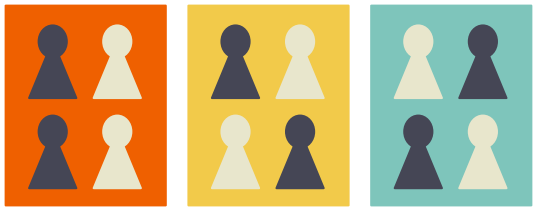


Stratified randomization

- **Stratified randomization:** group units of randomization into strata based on observable characteristics, and then randomize *within* strata

- **Typical Goal:** minimize variation between study arms: make sure that all strata are represented (typically, proportionally) in each arm
  - Strata with low internal outcome variance but high variance between them
  - Example: stratify by gender if (expected to be) strongly correlated with the outcome

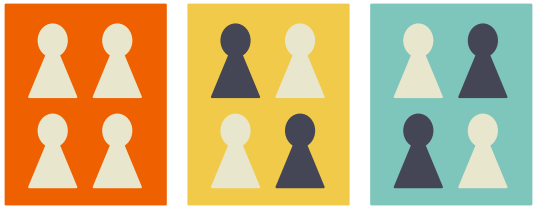
- Notes:
  - Must use a method to assign pre-defined shares of each stratum to treatment arms (often implicit!)
  - If combined with blocked randomization, the strata may be used as the blocks, ex. matched pair design.



Stratified randomization

- **Stratified randomization:** group units of randomization into strata based on observable characteristics, and then randomize *within* strata

- **Less frequent goal:** varying the probability of treatment by strata
  - Example: give some interventions only to a specific group or groups
  - In particular if an intervention depends on the presence of another



Stratified randomization: example

- Suppose there are 40 villages to be randomized into two study arms
- 10 of the villages have fewer than 1,000 residents

- **Coin flipping (“draw with replacement)** might result in imbalance in numbers: 25 villages in the control and 15 in the treatment arm
- **Fixed-share random assignment (“draw without replacement”)** ensures 20 villages in each arm, but might still result in unequal arms: e.g. 8 out of 10 small villages in the control

- **Stratification (with fixed-share assignment):** 5 of the small villages and 15 of the large villages into each arm → proportion of small villages in each arm matches overall proportion (25%)

Stratified randomization: notes

- Stratified randomization important for power if goal is to detect heterogeneous treatment effects
- Stratification (mostly) requires baseline information on the sample → **stratification variables**
  - Continuous variables (e.g., income) typically discretized
- Stratification applies at the **level of the intervention**: if two interventions are cross-randomized, they may have different units of randomization and/or different stratification variables
- **Estimation adjustments** to stratification:
  - Strata fixed effects, sometimes called LSDV
  - Weighted regression or weighted strata-specific TE or regression on strata dummies and strata-treatment interactions

Re-Randomization

- **Alternative to stratification:**
  - Randomize treatment assignment multiple times and then choose the “most balanced” sample

- Notes:
  - In extreme cases (e.g. stratification variable with two values such as gender) – equivalent to stratification
  - With continuous covariates: may effectively create correlated assignment probabilities (e.g. take case of four subjects with income levels *high, low, medium, medium* – high and low income will always be assigned to same study arm)

Other randomization methods

- **Covariate-adaptive randomization:**
  - Dynamically adjust probability of study arm assignment based on covariates of incoming subjects and already enrolled subjects

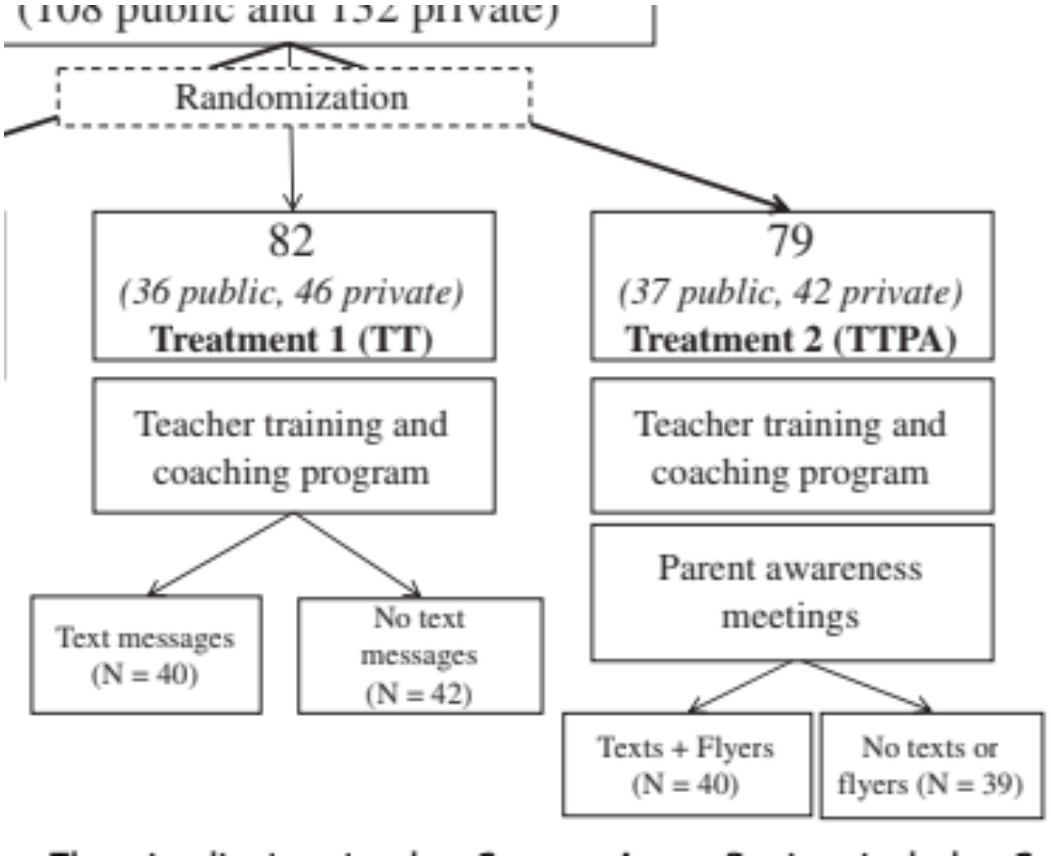
- **Other adaptive methods:**
  - Adjust probability of study arm assignment based on learning objectives (e.g. policy choice/best arm selection)

- Other: Anything else

Cross-Randomization

- **Cross-Randomization:** two or more interventions are randomly assigned
  - Often: study arms created for all combinations of interventions (*full* factorial design)
  - Sometimes: randomized assignment of one intervention stratified by random assignment of the other (*partial* factorial designs)

- Sometimes: interventions with different units of randomization
- **Random assignment in stages:**
  - A subset of units in all or some study arms receive an additional intervention
  - One sampling stage of the lower-level UoR may have “received intervention x” as an inclusion criteria



Example: TR 01

- Three study arms
- A subset of parents in the two treatment arms is randomized to receive texts
- *Stratification by treatment arm for the other intervention*

The choice of randomization approach impacts...

Empirical specification	Empirical specifications (Day 2)
Precision	Treatment effect estimates and prevision (Day 3)

3

“Adherence” to randomization

How is adherence to the randomization protocol reported?

Compliance

- **Compliance** is the adherence to randomized assignment.
    - Described e.g. as the percentage of a study arm that actually received the assigned treatment
- Real world challenges may result in imperfect compliance
- Non-compliance affects the ability to measure treatment effects.

Compliance impacts…

Empirical specification	Empirical specifications (Day 2)
Precision	Treatment effect estimates and prevision (day 3)

Balance

- A **balance test** may be conducted to examine how similar study arms are after randomization
- This often looks like regressing *baseline covariates* on *treatment assignment indicators*

Notes:

- Balance test regressions should take into account design choices such as stratified randomization
- Due to random variation, differences in covariates may still be statistically significantly different from zero with probability α
- Even if not: the two groups can still be very different!

Thank you for listening

Anja Sautmann
asautmann@worldbank.org

IDEAL
Impact Data and Evidence Aggregation Library