



Impact Data and Evidence Aggregation Library


# Fields review: Randomization units, stratification, and quality

Fei Yuan

*June 10, 2025*

## INTRODUCTION

**Fields in this section describe the units of randomization, stratification processes, and quality/robustness of the randomization**



Capturing this information correctly is required for generating the correct list of treatment effects





# Agenda

1. [Randomization units fields](#)
2. [Stratification fields](#)
3. [Quality and robustness fields](#)

1

# Randomization units

What-and how many-gets randomized



# Randomization unit fields

## Stage 1

1. Number of units of randomization
2. Units of randomization
3. Mapping units of randomization to interventions
4. Randomization method

## Stage 3

1. Number of randomization units *in study arm*
2. Number of randomization units *in study*



# Number of units of randomization

Description:

The total number of units of treatment assignment at the study level

Numeric value

*Multi-stage randomization if  $N > 1$ , also clustered*





# Unit of randomization

Description:

The unit of treatment assignment:

E.g., individuals, locations, facilities, groups, etc.

*Also referred to as the level of clustering, if there are lower-level units of analysis*





# Unit of randomization

Where to find:

- mostly found in the randomization or methods sections of a paper
- the text should indicate that the unit was used to assign the treatment
- study participant flow diagrams may reveal the units to which treatment was allocated







# Unit of randomization

## Coding instructions

- Select **all** units of randomization for the study one by one
- Unit of randomization is the level at which assignment of intervention to study arm were done
- Some units of randomization will generate follow up questions to specify further what the unit is
- Select 'other' and fill out the textbox if the unit of randomization does not exactly match the pre-specified units in the CV.
  - Other should be used sparingly!

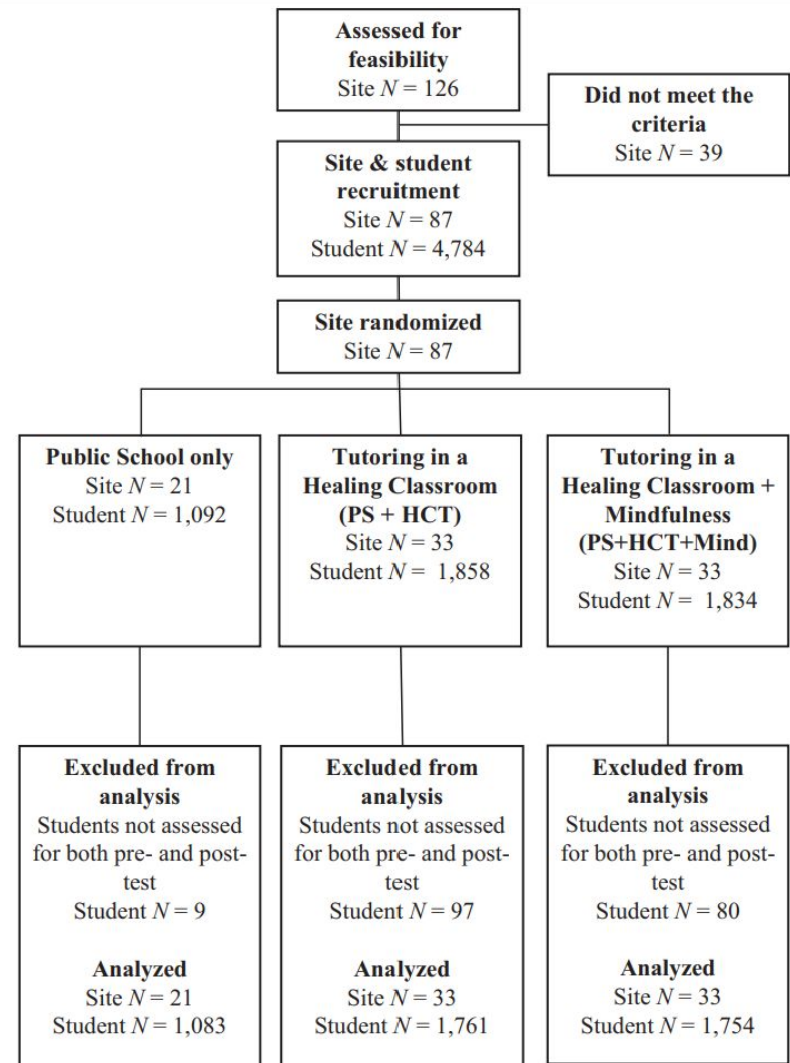


1. Individual	Any individual person, irrespective of demographic characteristics, professional, social or legal status, or affiliation.
1.1 Political/social leader	
1.2 Health provider	e.g. Doctors, nurses, midwives, etc.
1.3 Patient	
1.4 Education provider	e.g. Teachers, principals, etc.
1.5 Student	
1.6 Farmer	
1.7 Employee	
1.8 Business owner	
1.9 Voter	
1.10 Public servant	
1.11 Parent	
1.12 Other	
2. Organization or legal entity	Any kind of formal administrative and functional structure - includes associations, institutions, agencies, businesses, political parties, schools, etc.
2.1 Firm or business	
2.2 Legal or administrative division of a firm or business	e.g. Department
2.3 Farm or agricultural business	
2.4 School	
2.5 Legal or administrative division of a school	e.g. subjects, cohorts, grades
2.6 University/college	
2.7 Legal or administrative division of a university/college	e.g. majors, cohorts
2.8 Hospital, health clinic or doctor's office	
2.9 Other organization or legal entity	
3. Family	Two or more people related by blood, marriage (including step-relations), or adoption / fostering, or who identify as a couple, and who may or may not live together.
3.1 Nuclear family	
3.2 Extended family	
3.3 Parent(s) with dependent children	
3.4 Couples	
3.5 Other	
4. Household	A person or group of people who share common living arrangements or certain amenities, resources, or facilities. This may include pooling some or all of their income and wealth and collectively consuming certain types of goods and services, mainly housing and food.
5. Housing Unit	A house, apartment, mobile home, group of rooms, or single room that is occupied (or intended for occupancy) as separate living quarters in which the occupants live and eat separately from other building occupants.

specify

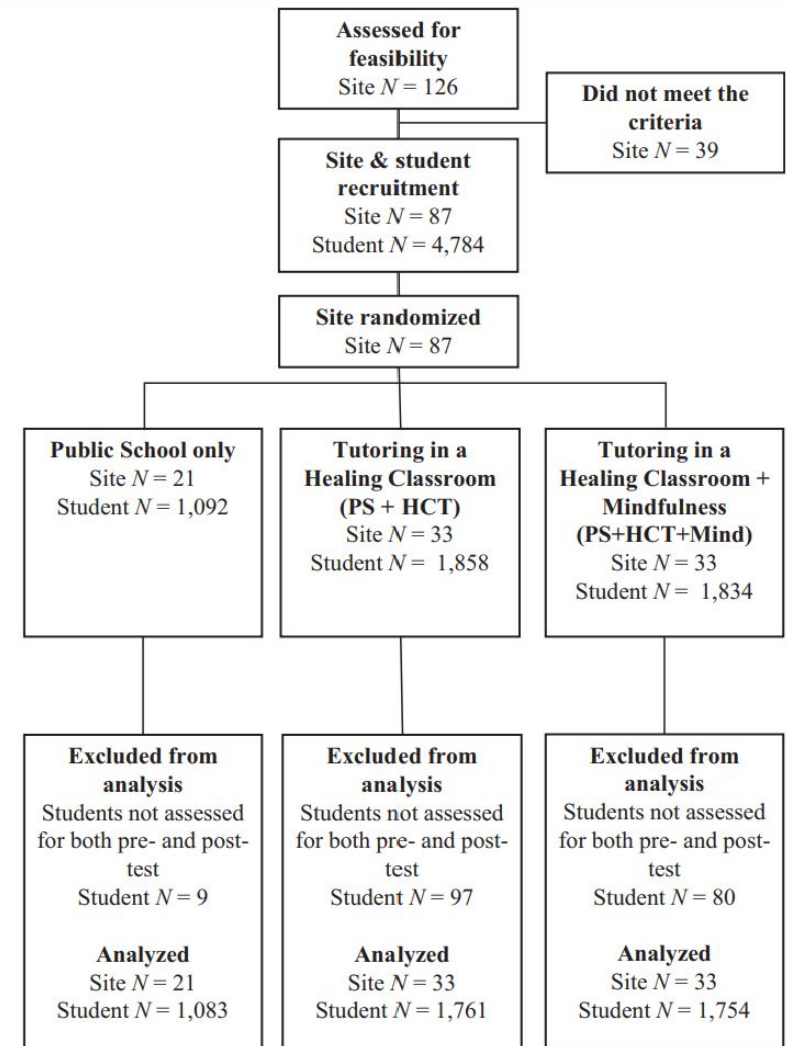
# Poll: what are *sites*?

- A. Units of randomization
- B. Units of analysis
- C. Both
- D. Neither



# Poll: what are *students*?

- A. Units of randomization
- B. Units of analysis
- C. Both
- D. Neither





# Unit of randomization

Unit of randomization = site

In CV, site = geographic location

Example from [CEGAD7](#)

- The 87 sites recruited in Akkar (n = 43) and Bekaa (n = 44) regions in Lebanon were stratified by region and randomized at the site level into one of the three treatment arms



# Mapping unit of randomization to interventions

Description:

For each of the interventions entered by the coder, one unit of randomization is assigned to the intervention

*Note: when there is only one unit of randomization, this question will be skipped*





# Mapping unit of randomization to interventions

Intervention 1: T1 -> community

Intervention 2: T2 -> community

The unit of randomization is the same between the two interventions

*Note: as this study has only one unit of randomization, the same unit of randomization must be used for all interventions and so this question would not be shown*

Example from [SIEFP10](#)

- Randomization allocated 35 communities per stratum of indigenous status to the intervention arms (T1: 934 children; T2: 754 children), and 32 communities per stratum to the comparison arm (T0: 784 children).

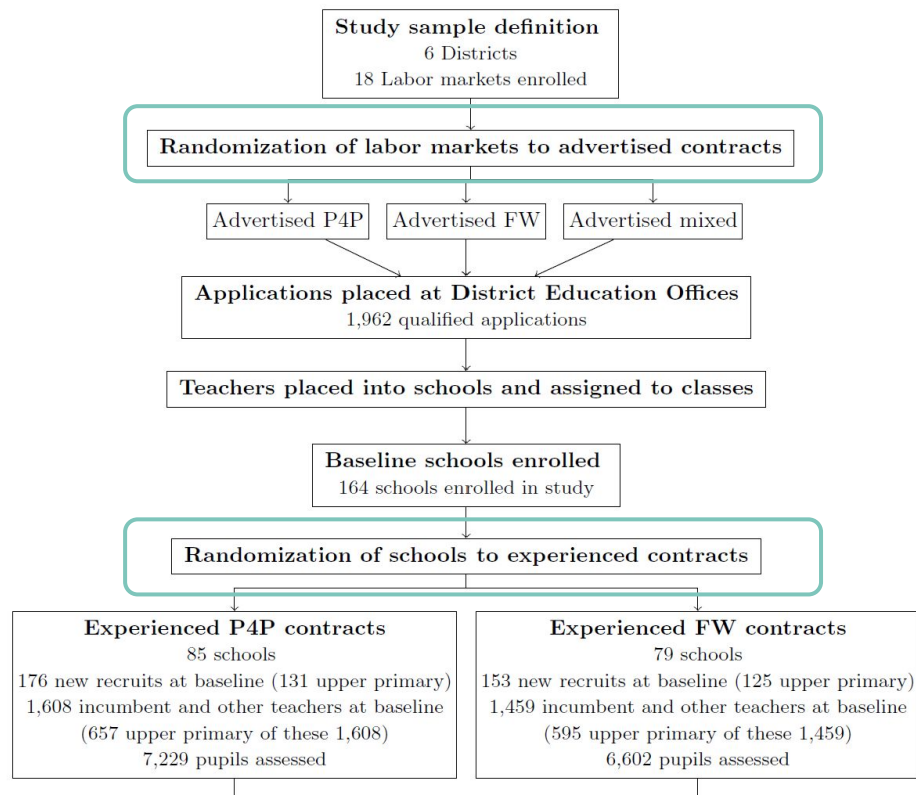
# Mapping unit of randomization to interventions

[Leaver et al. 2021](#)

Where to find:

- mostly found in the randomization or methods sections of a paper
- study participant flow diagrams may reveal the units to which treatment was allocated
- Flow diagrams often are found in supplementary materials

Figure A.1: Study profile







# Randomization method

Description:

Method used for randomization

*Select the method used by authors to randomly assign interventions to arms*





# Randomization method

CV, **select all**

- Simple randomization
- Fixed-share random assignment
- Block randomization/permutation blocking
- Stratification
- Matched pair design
- Re-randomization
- Covariate adaptive randomization
- Other

*Other should be used sparingly, and will generate a text box asking you to describe the randomization procedure.*



# Randomization method

**Simple randomization** is based on a single sequence of random assignment, e.g. flipping a coin or drawing with replacement.

- Drawing a treatment arm assignment for each unit of randomization with fixed probabilities.

**Fixed-share random assignment** is based on a single share of random assignment, e.g. drawing without replacement.

- Selecting a random subset of pre-selected size into each arm, e.g. randomly 20 villages to each arm.





# Randomization method

**Block randomization/permutation blocking** is intended to equalize the number of units per treatment arm

*Note: In most papers, block assignment will appear in the description of the randomization procedure and will use phrases such as “blocked assignment” or “treatment within blocks.” Block randomization is intended to equalize the number of units per treatment arm*





# Randomization method

**Stratification** is intended to equalize characteristics of the treatment arms

*Note: In most papers, stratification will appear in the description of the randomization procedure and will use phrases such as "stratified" or "treatment randomly assigned within...". Note that strata can be used in both sampling(how units came to be in the study) and randomization (how units came to be in the study arms). Be careful to mark stratified randomization in this field ONLY if the paper does stratified randomization*





# Randomization method

**Matched pair design** is stratification with two units per stratum/block.

**Covariate adaptive randomization** assigns participants to treatment and control group by taking into account the specific covariates and previous assignments of participants. It often uses the method of minimization to reduce the imbalance of sample size among several covariates.

**Re-randomization** assigns consists of redoing the randomization until some pre-specified balance criterion on the observed covariates is met.



## Units of randomization (Stage 1)

The **types** of unit over which randomization occurred

If interventions were randomized across schools and principals, there are **two units of randomization**: schools, principals

## Randomization units (Stage 3)

The **distinct** units in the study

If there are 300 schools and 300 principals in the study, there are **600 randomization units** (300 schools, 300 principals)



# Number of randomization units *in study arm*

Description:

The total number of randomization units in each study arm

*Note: this question will be repeated for each arm of the study (including the control arm) by unit of randomization.*







# Number of randomization units *in study arm*

Where to find:

- mostly found in the research design sections of the paper
  - specifically in the description of the random assignment,
- may also be found in participant flow diagrams (e.g. the CONSORT flow diagram) or
- a table that disaggregates information by treatment arms (e.g., a balance table)





# Number of randomization units *in study arm*

Arm 1: control–21 sites

Arm 2: public school + tutoring in a healing classroom–33 sites

Arm 3: public school + tutoring in a healing classroom + mindfulness–33 sites

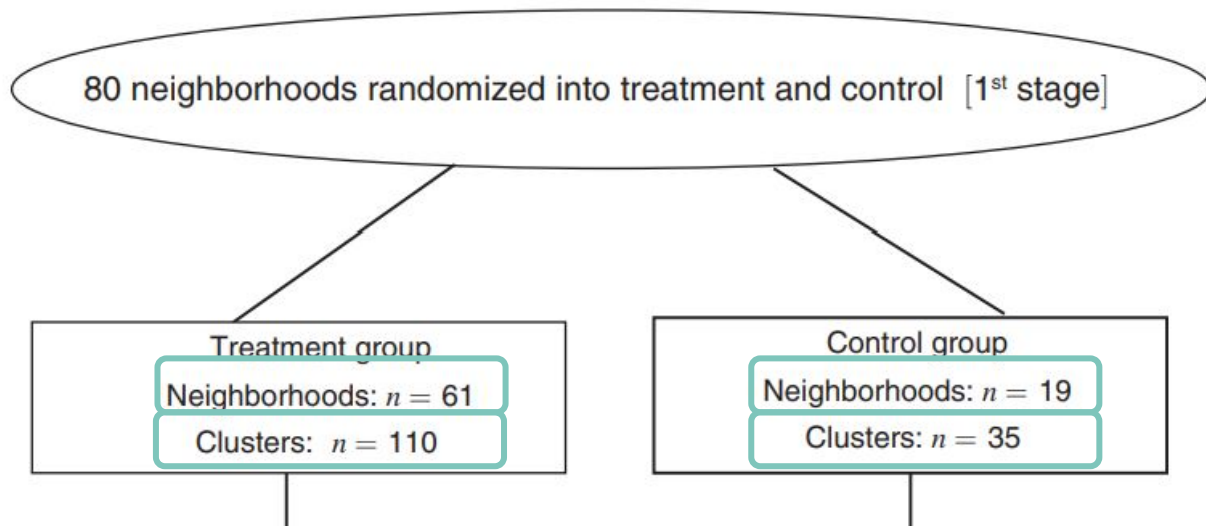
Example from [CEGAD7](#)

- The 87 sites... were stratified by region and randomized...into one of the three treatment arms:
- (1) 21 control sites
- (2) 33 public school + Tutoring in a Healing Classroom sites
- (3) 33 public school + Tutoring in a Healing Classroom + Mindfulness sites



# Number of randomization units *in study arm*

Two-stage  
randomization:  
reports number of  
randomization  
units for both.





# Number of randomization units *in study*

## Description:

The total number of randomization units in the study

## Coding instructions

- If it is not possible to ascertain the number of randomization units for **any of the distinct arms** in the study, enter “-99”
- You will then be asked about the total number of randomization units *in the study*

*Note: this question will be asked only if it is not available at the study arm level*

2

# Stratification

How is the randomization stratified?



# Stratification fields

Randomization method, **select all**

- Simple randomization
- Fixed-share random assignment
- Block randomization/permutation blocking
- **Stratification**
- **Matched pair design**
- Re-randomization
- Covariate adaptive randomization
- Other



## Stage 1

1. Number of stratification variables
2. Stratification variables
3. Stratification for study arms
4. Mapping of stratification variables to interventions



# Number of stratification variables

Description:

Number of variables used to create strata

*Note: Number of stratification variables is the number of distinct **variables** used in the randomization process; not the number of distinct values of those variables*





# Stratification variables

Description:

A description of the variable used in stratification

*Note: This question repeats for the number of times entered by the coder in the “number of stratification variables” question*







# Stratification variables

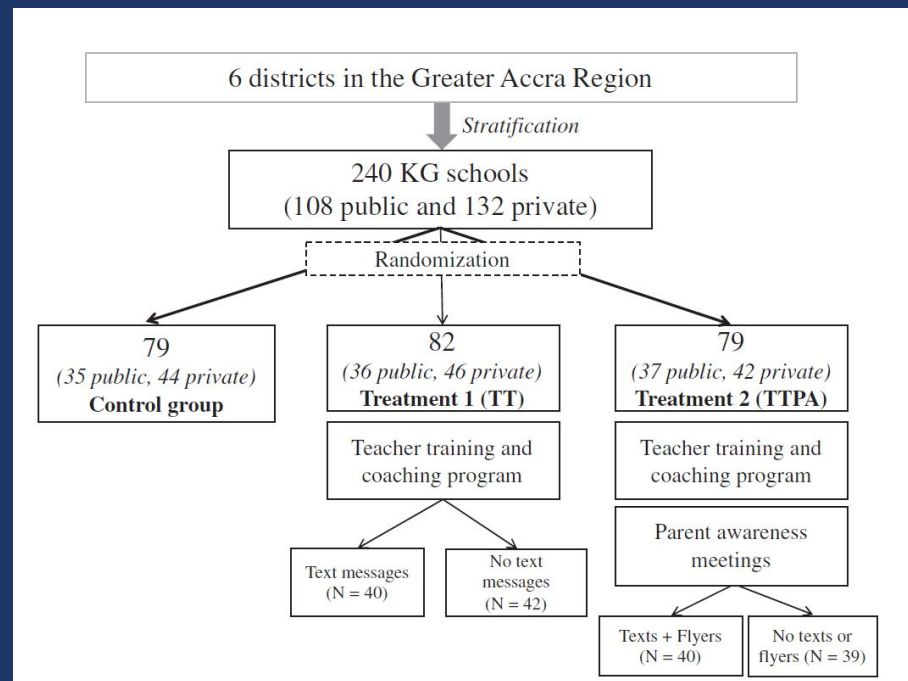
Coding instructions:

- Report these variables in the order they appear in the paper
- If higher-level interventions are the strata, report an indicator for being in intervention X
- If local terms are used (e.g., panchayat), retain local term rather than English approximation (e.g., village council)



# Poll: in TR01, what are the stratification variables?

- A. Public/private school
- B. Gender of school principal
- C. Region of country
- D. Indicator for being assigned to the TT group
- E. Indicator for being assigned to the TTPA group



**Figure 2.** Research design. *Note.* The six districts in the Greater Accra Region include: Ga South, Adenta, Ledzokuku-Krowor, Ga Central, La Nkwantanang-Madina, and Ga West. School randomization is stratified by district and public- versus private-sector status.



# Stratification variables

Variable 1: district

Variable 2: public/private sector

Variable 3: Indicator for being in the TT intervention

Variable 4: Indicator for being in the TTPA intervention

## Example from [TR01](#)

- First, randomization was conducted at the school level and stratified by district and public/private sector ....  
Second, half of the TT treatment schools were randomly assigned to receive weekly text messages for teachers and half of the TTPA treatment schools were randomly assigned to receive teacher text messages



# Stratification for study arms

Description:

Report if the same (set of) stratification variables are used for assigning all interventions for study arms

*Note: If the answer is yes, you will not need to map the stratification variables to study arms*





# Mapping stratification variables to interventions

Description:

For each of the interventions entered by the coder, enter the stratification variables used to randomize that intervention to randomization units





# Mapping stratification variables to interventions

## Intervention 1: TT

- District
- public/private sector

## Intervention 2: TTPA

- District
- public/private sector

## Intervention 3: Text messages to teachers

- Indicator for being in the TT intervention
- Indicator for being in the TTPA intervention

### Example from [TR01](#)

- First, randomization was conducted at the school level and stratified by district and public/private sector .... Second, half of the TT treatment schools were randomly assigned to receive weekly text messages for teachers and half of the TTPA treatment schools were randomly assigned to receive teacher text messages

## 2 Quality and robustness

Was the randomization procedure followed?



# Quality and robustness fields

Stage 3

1. Compliance
2. Balance test





# Compliance

Description:

Report the fraction of the study arm that was correctly assigned the interventions mapped to the study arm

*Note: Repeat for each arm in the study. If this cannot be reported, enter “-88” and you will be shown an open-text field to report more details*





# Compliance

Hints:

Compliance or take-up?

Recall: **take-up** is the share of an arm that took up an intervention. For arms assigned to an intervention, take-up is **compliance**. For arms not assigned to an intervention, take-up is **non-compliance**





# Compliance

Compliance or take-up?

Suppose there is one intervention: a cash transfer, randomized at the village level. How do we measure compliance?





# Compliance

Compliance, take-up, and implementation fidelity?

		Village was randomly assigned...	
		<b>To <i>not</i> receive</b> the cash transfer (70 villages)	<b>To receive</b> the cash transfer (70 villages)
Village in real world	<b>Did not</b> receive cash transfer	60 villages (compliance)	10 villages (not offered) (Implementation fidelity)
	<b>Refused</b> cash transfer		10 villages (non-compliance/Didn't take up)
	<b>Did</b> receive cash transfer	10 villages (non-compliance/take-up)	50 villages (compliance/take-up)



# Compliance

This is take-up—villages that *actually took the treatment, regardless of their treatment status*

		Village was randomly assigned...	
		To <i>not</i> receive the cash transfer	To receive the cash transfer
Village...	<b>Did not</b> receive cash transfer	60 villages	10 villages
	<b>Did</b> receive cash transfer	5 villages	70 villages



# Compliance

This is compliance – villages that followed what they were randomly assigned to do

		Village was randomly assigned...	
		To <i>not</i> receive the cash transfer	To receive the cash transfer
Village...	<b>Did not</b> receive cash transfer	60 villages	10 villages
	<b>Did</b> receive cash transfer	5 villages	70 villages



# Compliance

Note that take-up is compliance **for arms assigned to the intervention**; but *non-compliance* for arms not assigned to the intervention

		Village was randomly assigned...	
		To <i>not</i> receive the cash transfer	To receive the cash transfer
Village...	<b>Did not</b> receive cash transfer	60 villages	10 villages
	<b>Did</b> receive cash transfer	5 villages	70 villages



# Compliance

With multiple interventions

		Village was randomly assigned...		
		To receive <b>no</b> interventions	To receive intervention A	To receive intervention B
Village...	<b>Received no interventions</b>	60 villages	10 villages	5 villages
	<b>Received intervention A</b>	5 villages	70 villages	5 villages
	<b>Received intervention B</b>	10 villages	5 villages	40 villages





# Compliance

## Take-up of intervention A

		Village was randomly assigned...		
		To receive <b>no</b> interventions	To receive intervention A	To receive intervention B
Village...	<b>Received no interventions</b>	60 villages	10 villages	5 villages
	<b>Received intervention A</b>	5 villages	70 villages	5 villages
	<b>Received intervention B</b>	10 villages	5 villages	40 villages



# Compliance

## Take-up of intervention B

		Village was randomly assigned...		
		To receive <b>no</b> interventions	To receive intervention A	To receive intervention B
Village...	<b>Received no interventions</b>	60 villages	10 villages	5 villages
	<b>Received intervention A</b>	5 villages	70 villages	5 villages
	<b>Received intervention B</b>	10 villages	5 villages	40 villages



# Compliance

## Compliance

		Village was randomly assigned...		
		To receive <b>no</b> interventions	To receive intervention A	To receive intervention B
Village...	<b>Received no interventions</b>	60 villages	10 villages	5 villages
	<b>Received intervention A</b>	5 villages	70 villages	5 villages
	<b>Received intervention B</b>	10 villages	5 villages	40 villages



# Compliance

Hints:

- Non-compliance should only cover cases where a mistake in randomization lead to treatment being offered to and taken by the wrong units.
- In other cases, report that this information is not available





# Compliance

## Coding instructions

- Enter fraction in this field
  - For example, 15/16 indicating 15 out of 15 randomization units complied with the treatment status or 49/100 for a compliance rate of 49%.
- Enter “-99” if the information is not mentioned in the paper.
- Enter “-88” if compliance rate cannot be entered as numeric values. Please specify the details.





## Poll: What is compliance rate for the **control group**?

- A. 49%
- B. 97%-98%
- C. 2%-3%
- D. Unknown

Example from Bos et al. 2024

Table 6 shows that due to imperfect compliance, the differential likelihood of receipt of the child development card, household picture book, and nature picture book between treatment and control households was approximately 49 percentage points (instead of 100 under perfect compliance). Furthermore, 2%–3% of households in the control group received these materials."



## Poll: What is compliance rate for the **treatment group**?

- A. 49%
- B. 97%-98%
- C. 2%-3%
- D. Unknown

Example from Bos et al. 2024

Table 6 shows that due to imperfect compliance, the differential likelihood of receipt of the child development card, household picture book, and nature picture book between treatment and control households was approximately 49 percentage points (instead of 100 under perfect compliance). Furthermore, 2%–3% of households in the control group received these materials."



# Balance table

Description:

Indicates whether the paper includes a balance test, in the main paper or supplementary materials.

- If yes, report the table number





Fei Yuan  
fyuan@worldbank.org

Thank you  
for listening



Impact Data and Evidence Aggregation Library