IDEAL

Impact Data and Evidence Aggregation Library

Session: RCT Design

Jennie Barker

June 9, 2025

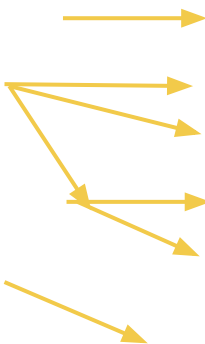## This session discusses the fields related to RCT Design

These fields ask structural questions about each RCT that other parts of the survey build upon in Stage 2 and 3.

Agenda

IDEAL Concept Review

1. Experiment vs. intervention
2. Intervention assignment strategies
3. Study arms
4. Rounds of data collection

Fields

1. Number of experiments
2. Intervention assignment strategy
3. Number of interventions
4. Total number of study arms
5. Mapping interventions to study arms
6. Number of rounds of data collection (study level)

I. Number of experiments

- **Definition:** Number of experiments under evaluation in the paper
  - Study population and unit of randomization
  - Same set of interventions
  - Randomization used to create **comparable** treatment arms

I. Number of experiments

- Most assigned papers we will code have **one field experiment**
  - If there are **0 experiments**, flag to a supervisor (This will be rare!)
    - If the paper includes a **lab-in-the field experiment** or an **experiment with design** rather than policy interventions, flag to a supervisor

  - If there are **multiple experiments**, you will be prompted to repeat the Stage 1 survey for each experiment
    - Papers with results across **multiple countries** usually have multiple experiments
    - There are occasionally papers that will report results from two experiments in the same country (e.g., drawing on **different study populations)**

Example: Multiple Experiments

Barrera-Osorio et al. (2011)

- This paper includes 2 experiments
  - Different study populations

B. Structure of Randomization

As required by the SED, the assessment of the treatments was divided into two separate experiments located in two very similar localities in Bogota, San Cristobal, and Suba.[7] Eligible registrants in San Cristobal, ranging from grade 6–11, were randomly assigned between a control group, the basic treatment, and the savings treatment. The tertiary treatment was evaluated separately in an experiment in Suba, where students ranging from grade nine through eleven, were randomly assigned to either the tertiary treatment or a control group. Hence, while we can experimentally estimate the causal effect of the tertiary treatment, we can only identify its relative effect compared to the basic (or saving treatment) by relying on a rich set of socio demographic controls and school-level fixed effects rather than purely random variation.[8]

Example: Multiple experiments

We selected participants for the endline intervention from among the 450 control group respondents slated to provide a day-long LENA recording at endline (see Section II.C for details on how respondents were selected for the LENA subsample). This sampling frame allows us to use their first LENA recording as a pre-endline-intervention measure of the child's auditory environment. For budget reasons, we chose 225 of these women for the endline intervention, randomly selecting them with stratification by child age and baseline self-reports of behavior.[14]

Dupas et al. 2024

- This paper features an intervention that is conducted at endline among some respondents in the control group. It thus has two experiments.

Example: Design intervention

Jeong et al. (2023)

- This paper evaluates a design intervention, in which survey questions are randomized to examine survey fatigue, rather than a policy intervention.
- This paper would have **0** experiments for our pilot.

2.2. Question order randomization

This experiment takes place within baseline and endline surveys which are similar to the World Bank's LSMS surveys and take about 2.5 h to complete on average. The surveys contain 19 self-contained sections, including household demographics, agriculture, income, expenditures, savings, assets, labor supply, shocks, and other topics. We show the full list of sections in Fig. 1.

The beginning of the survey (which included household identifying information, demographics, and agriculture) and the end of the survey (which had a section on intimate partner violence, followed by the collection of household tracking information) were the same across all versions. The remaining sections were grouped into 3 modules, and the order of these 3 modules was randomized, giving us 6 versions of the survey (which we refer to as versions A–F — see Fig. 2). The survey software records the amount of time elapsed (since beginning) at each question, allowing us to calculate the exact time at which a question appeared in the survey.

Example: Lab-in-the-field experiment

De Martino et al. (2015)

- This paper evaluates a lab-in-the-field experiment, in which participants were made hypothetical offers.
  - This paper would have **0** experiments for our pilot.

Data and Experimental Setup

Data Collection

We ran a lab-in-the-field experiment within the baseline survey for a pilot PES program in the state of SP. The baseline survey was carried out in 2013 in two municipalities, Ibiúna and Guapiara, both located in the southeast region of SP state. Each municipality established priority areas for conservation

Number of experiments check

- After you review the number of experiments and enter the number, you will receive a question asking you to confirm the number if you enter **0** or more than **1.**
  - **0:** confirm that there are no eligible experiments
  - **>1:** confirm that there is more than one eligible experiment

Country

- **Definition:** The country in which the experiment is implemented
  - Even if the experiment only takes place in a village or region of a given country
    - Title, Abstract, Introduction
  - You will be asked for the **ISO code** (from a dropdown menu)
- Examples
  - Alatas et al. 2012: Indonesia (**IDN**)
  - Chong et al. 2015: Mexico (**MEX**)
  - Wolf et al. 2019: Ghana (**GHA**)

Subnational Location

- **Definition:** The subnational location where the experiment took place
  - **Largest** geographic location(s) within a country in which the experiment took place
    - Open text on city, state, or region
  - For the specific experiment in question
- Examples
  - Alatas et al. 2012: North Sumatra, South Sulawesi, and Central Java (provinces)
  - Chong et al. 2015: Jalisco, Morelos, Tabasco (states)
  - Wolf et al. 2019: Greater Accra Region (region)

II. Intervention Assignment Strategy

- **Definition:** The strategy used for assigning interventions to study arms
- In the pilot, we code papers with the following designs:
  - **Parallel**: Each intervention is assigned to only one arm
  - **Factorial:** At least one intervention is assigned to more than one study arm
  - **Crossover:** Each study arm receives different intervention (including no intervention in different phases of the study
    - *If study endline occurs **prior to** study units receiving the interventions beyond what they were initially assigned to, do not select this option*

II. Intervention Assignment Strategy

- There are other assignment strategies, which we will not code in the pilot:
  - **Adaptive:** The rule by which interventions are assigned can change throughout the course of the study

Parallel Design

Crawfurd et al. (2023)

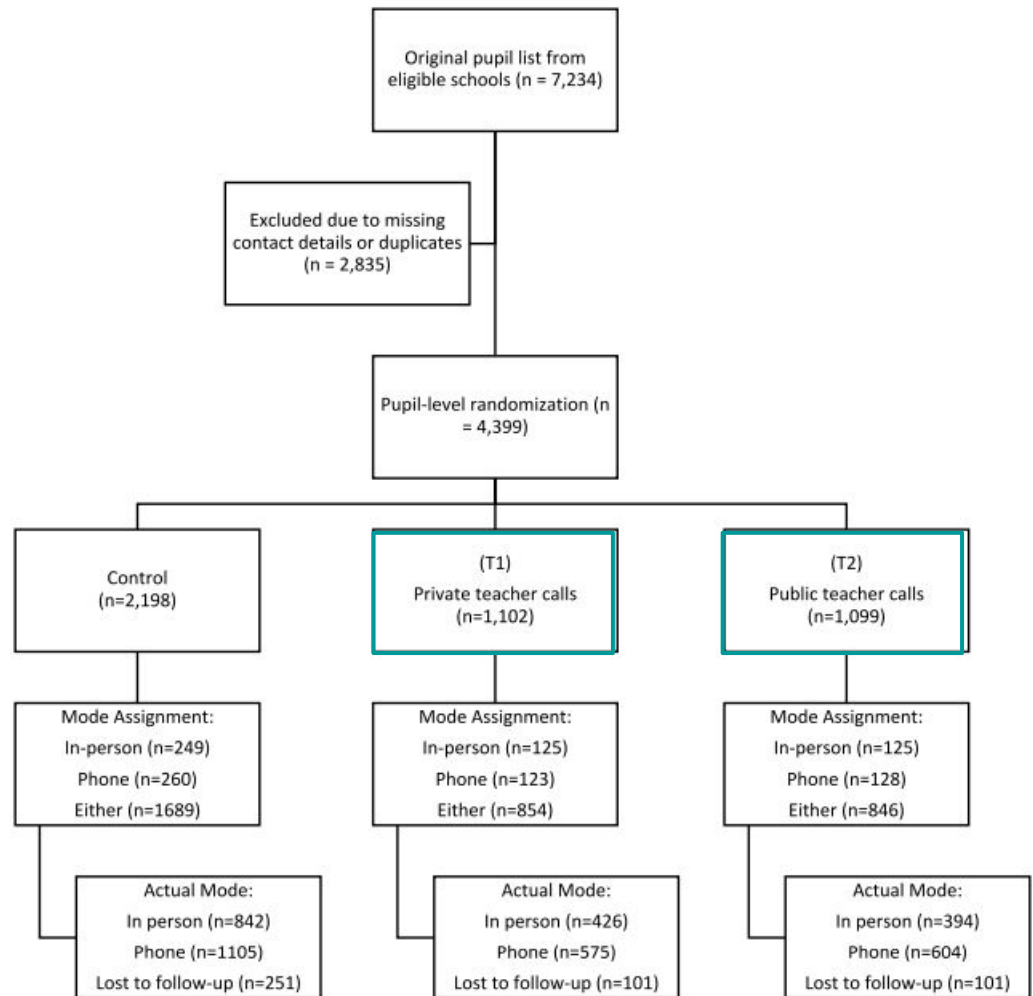- Each treatment arm is assigned a unique intervention

Original pupil list from eligible schools (n = 7,234)

Excluded due to missing contact details or duplicates (n = 2,835)

Pupil-level randomization (n = 4,399)

| Control (n=2,198) | (T1) Private teacher calls (n=1,102) | (T2) Public teacher calls (n=1,099) |

Mode Assignment:
In-person (n=249)
Phone (n=260)
Either (n=1689)

Mode Assignment:
In-person (n=125)
Phone (n=123)
Either (n=854)

Mode Assignment:
In-person (n=125)
Phone (n=128)
Either (n=846)

Actual Mode:
In person (n=842)
Phone (n=1105)
Lost to follow-up (n=251)

Actual Mode:
In person (n=426)
Phone (n=575)
Lost to follow-up (n=101)

Actual Mode:
In person (n=394)
Phone (n=604)
Lost to follow-up (n=101)

Fig. 1. Consort diagram

Factorial Design

Wolf et al. (2019)

- Treatment arms receive some of the same interventions
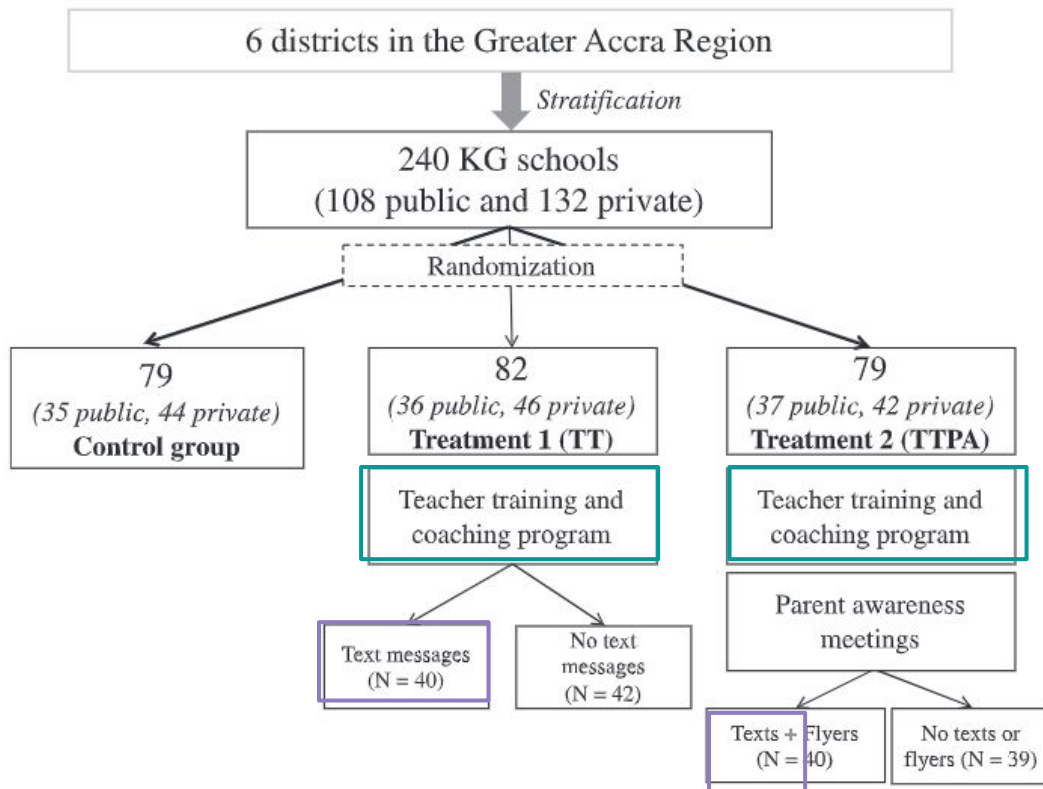
Crossover design

Miguel and Kremer (2004)

- Vary the timing in which the different groups receive the treatment
  - A comparison group in 1998 receives the intervention in 1999

school was assigned to a given project group.[9] Due to ICS's administrative and financial constraints, the health intervention was phased in over several years. Group 1 schools received free deworming treatment in both 1998 and 1999, Group 2 schools in 1999, while Group 3 schools began receiving treatment in 2001. Thus in 1998, Group 1 schools were treatment schools, while Group 2 and Group 3 schools were comparison schools, and in 1999, Group 1 and Group 2 schools were treatment schools and Group 3 schools were comparison schools.

Poll

Dolan et al. (2022): Which design?

A. Parallel
B. Factorial
C. Crossover
D. Not sure - help!
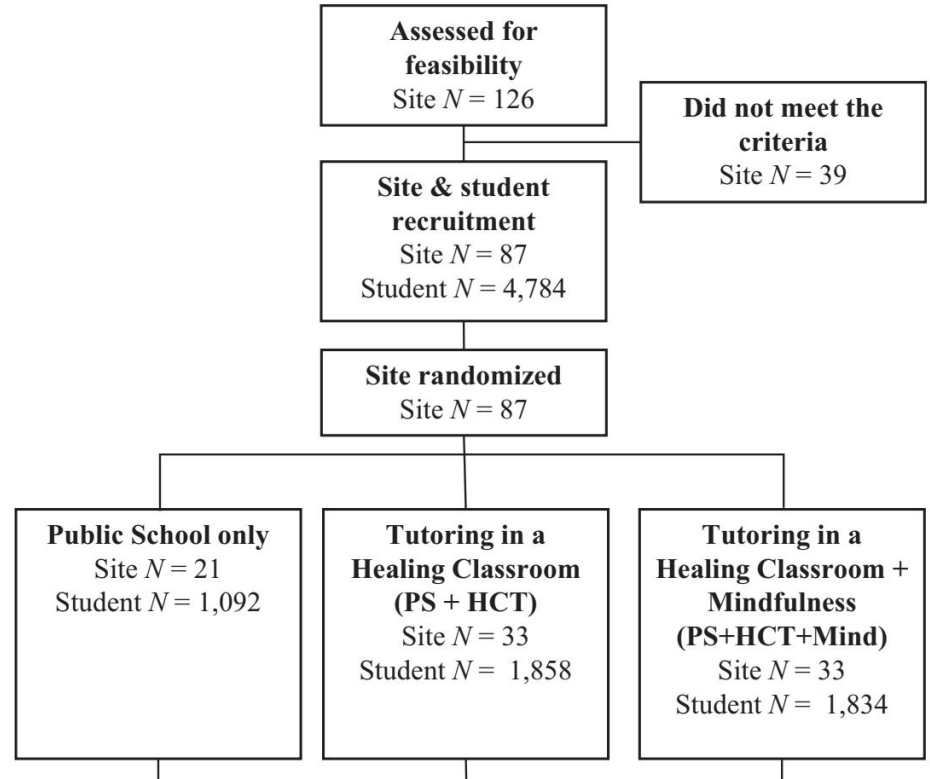
Poll

Lopez et al. (2022) Which design?

A. Parallel
B. Factorial
C. Crossover
D. Not sure - help!

of the rainy season and therefore the period of highest malaria risk. We divided the 60 clinics into 3 cohorts of 20 based on geography. Each of the three cohorts rotated through two weeks of data collection and experimental intervention. Within each cohort, we randomly assigned each clinic to 1 of 20 intervention schedules depicted in Figure 2. Each clinic received two Control days, two Doctor Voucher days, and two Patient Voucher days. Although all clinics were informed of the upcoming study

III. Number of interventions

- **Definition:** Number of distinct interventions in the study
  - Include and label any intervention **beyond the status quo** that is administered to the comparison/control group
  - Interventions with different intensity levels should be **counted separately**
    - E.g., Low vs. high intensity training program

Wolf et al. (2019) has 4 distinct interventions, 2 of which are common
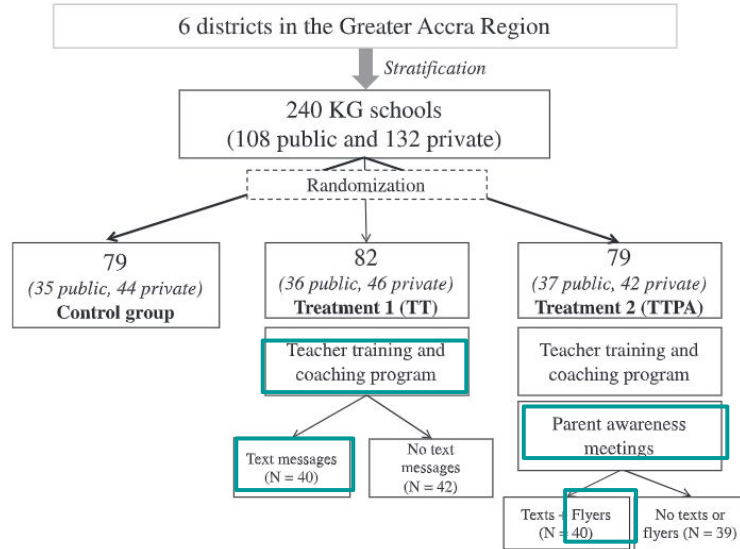
Barnhart et al. (2020) has 2 distinct interventions, one of which is common

1806

District. Families were visited by study staff to confirm eligibility for the study and enrollment in VUP. Because a single CBV is expected to provide Sugira Muryango to up to five families, enumerated families were grouped into geographic clusters which included at least five families that could feasibly be served by a single coach. To reach our target sample of 40 families, we selected eight clusters to participate in the trial and randomized each cluster to either the intervention or VUP cash-only control arm. In each cluster, five families were randomly selected for enrollment in the study. If, at enrollment, families were no longer eligible for the program or interested in participating, a replacement family was randomly selected from the remaining families in that cluster until no more families

Chong et al. (2015): 3 interventions

In total, 150 voting precincts were randomly assigned to each of the three interventions, for a total of 450 treatment and 1,910 control precincts. We distributed 44,000 flyers

1. Corruption information
2. Budget information
3. Poverty expenditure
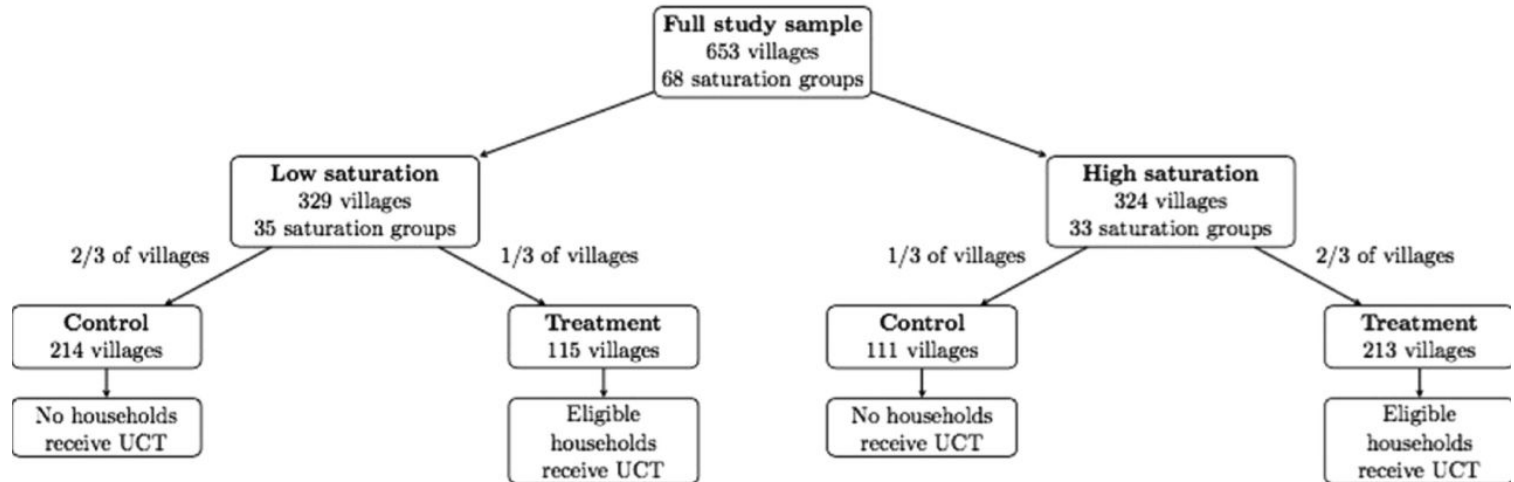
Miguel and Kremer (2004): 3 interventions

1. Free deworming treatment (program intervention)
2. Treatment in 1998 (timing intervention)
3. Treatment in 1999 (timing intervention)

Poll

Egger et al (2022): How many interventions?

A. 1
B. 2
C. 3
D. 4

(a) Randomization

Full study sample
653 villages
68 saturation groups

Low saturation
329 villages
35 saturation groups

High saturation
324 villages
33 saturation groups

2/3 of villages   1/3 of villages   1/3 of villages   2/3 of villages

Control
214 villages

Treatment
115 villages

Control
111 villages

Treatment
213 villages

No households
receive UCT

Eligible
households
receive UCT

No households
receive UCT

Eligible
households
receive UCT

Intervention label

- A short, <u>author-used</u> label for the intervention
    - What is used in tables, figures, or participant flow diagrams

6 districts in the Greater Accra Region

Stratification

240 KG schools
(108 public and 132 private)

Randomization

79
(35 public, 44 private)
Control group

82
(36 public, 46 private)
Treatment 1 (TT)

79
(37 public, 42 private)
Treatment 2 (TTPA)

Teacher training and coaching program

Teacher training and coaching program

Text messages (N = 40)

No text messages (N = 42)

Parent awareness meetings

Texts + Flyers (N = 40)

No texts or flyers (N = 39)

IV. Total Number of Study Arms

- **Definition:** The number of study arms, i.e., subgroups of participants that receive none, one, or several specific interventions
  - Including control group
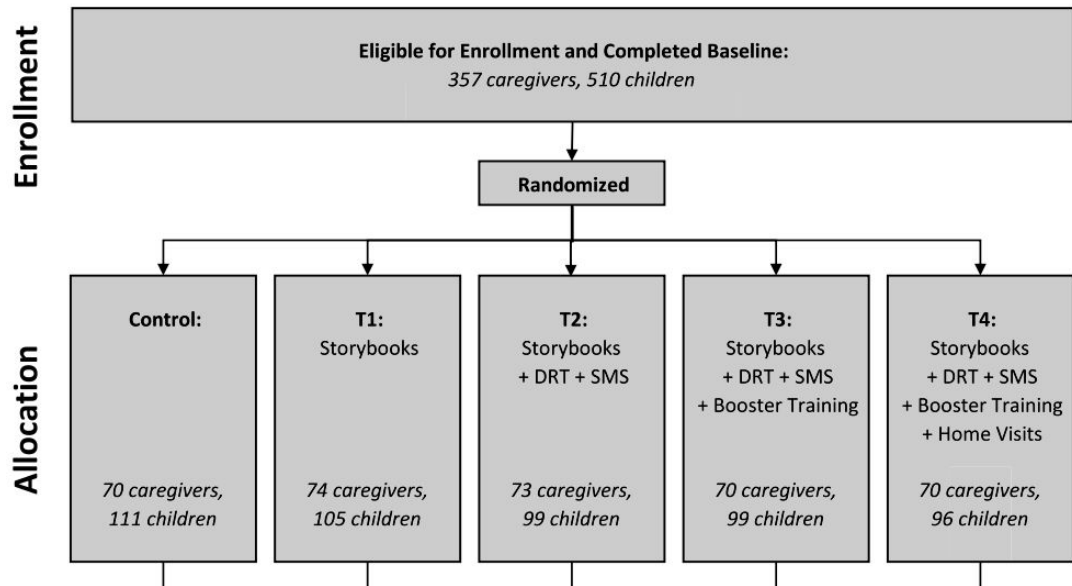 - Subgroup: receives same (set of) intervention(s)

Example

Moehler and Conroy-Krutz
(2016) have 4 study arms

Subjects were randomly assigned to photograph or non-photograph conditions. We also designed the experiment to allow us to test for the effects of party cues on party-based voting; those results are reported separately in Conroy-Krutz et al. (2015). Therefore, some ballots in each of these conditions included party identifiers (i.e., names and symbols), while others did not, resulting
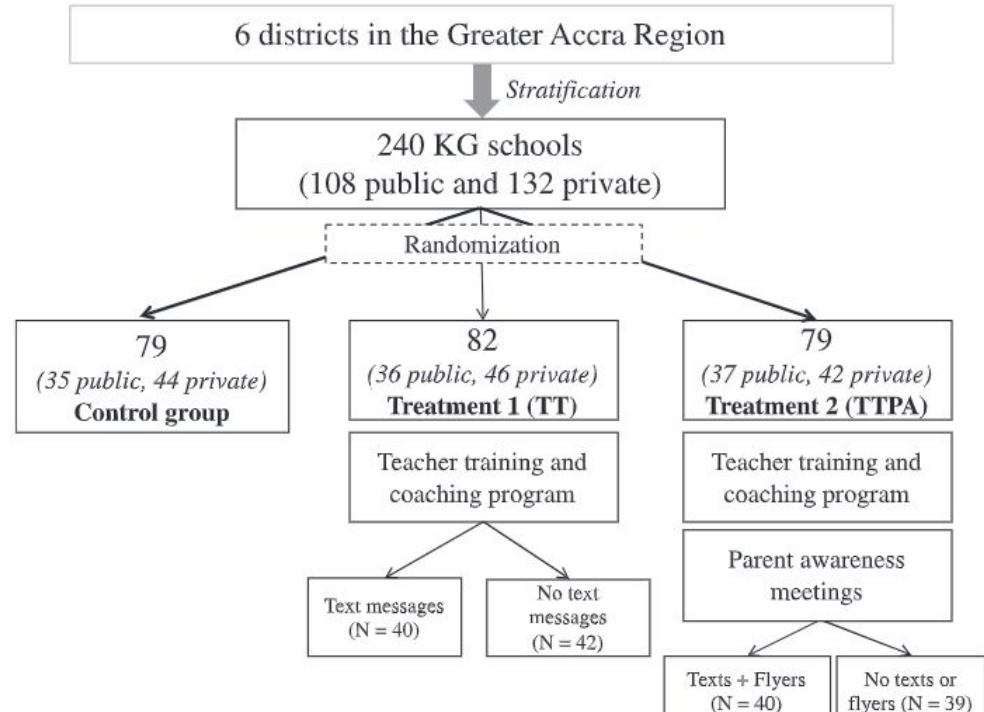
Example

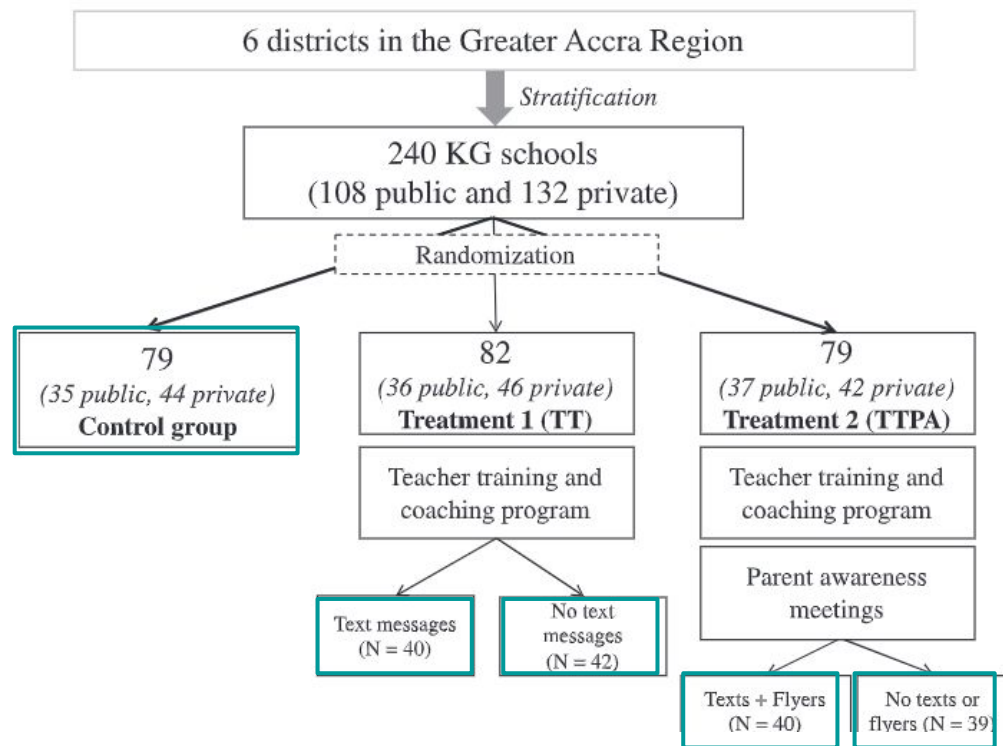Knauer et al. (2020) have 5
study arms

Poll

Wolf et al. (2019): How many arms?

A. 3
B. 4
C. 5
D. 6

6 districts in the Greater Accra Region

Stratification

240 KG schools
(108 public and 132 private)

Randomization

79
(35 public, 44 private)
Control group

82
(36 public, 46 private)
Treatment 1 (TT)

79
(37 public, 42 private)
Treatment 2 (TTPA)

Teacher training and coaching program

Teacher training and coaching program

Text messages (N = 40)

No text messages (N = 42)

Parent awareness meetings

Texts + Flyers (N = 40)

No texts or flyers (N = 39)

6 districts in the Greater Accra Region

Stratification

240 KG schools
(108 public and 132 private)

Randomization

79
(35 public, 44 private)
Control group

82
(36 public, 46 private)
Treatment 1 (TT)

Teacher training and coaching program

Text messages
(N = 40)

No text messages
(N = 42)

79
(37 public, 42 private)
Treatment 2 (TTPA)

Teacher training and coaching program

Parent awareness meetings

Texts + Flyers
(N = 40)

No texts or flyers (N = 39)

V. Mapping Interventions to Study Arms

- **Definition:** Intervention(s) received by each study arm
  - Start with control arm
    - If control arm received status quo, select "None"
  - For factorial designs, be sure to include **all interventions** received by each arm

Example

Subjects were randomly assigned to photograph or non-photograph conditions. We also designed the experiment to allow us to test for the effects of party cues on party-based voting; those results are reported separately in Conroy-Krutz et al. (2015). Therefore, some ballots in each of these conditions included party identifiers (i.e., names and symbols), while others did not, resulting
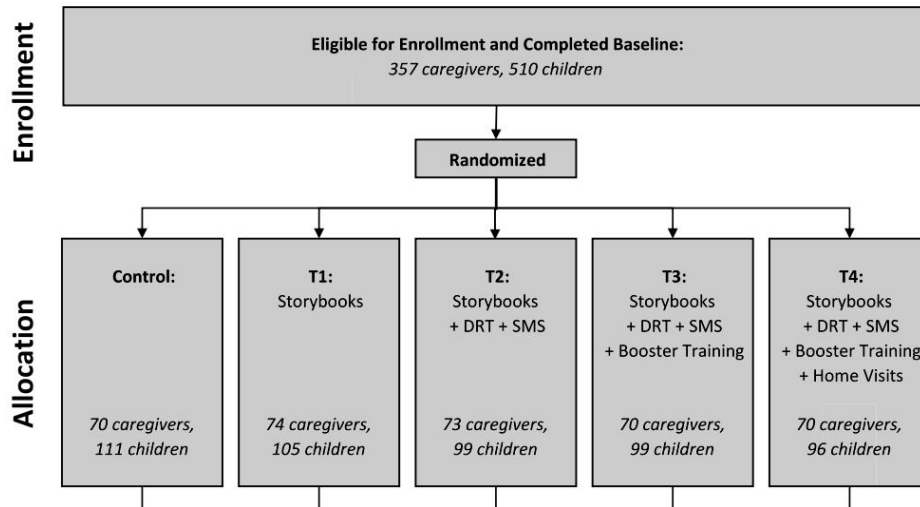
- Moehler and Conroy-Krutz (2016) have 4 study arms

Arm 1: Non-photograph ballot + None

Arm 2: Non-photograph ballot + Party identifiers (names and symbols)

Arm 3: Photograph ballot + None

Arm 4: Non-photograph ballot + Party identifiers (names and symbols)

Example

- Knauer et al. (2020) have 5 study arms

Arm 1: None
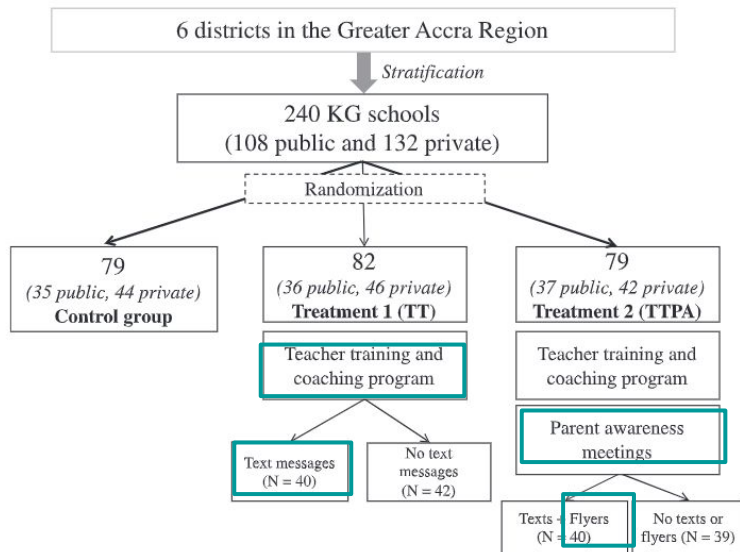
Arm 2: Storybooks

Arm 3: Storybooks + DRT + SMS

Arm 4: Storybooks + DRT + SMS + Booster Training

Arm 5: Storybooks + DRT + SMS + Booster Training + Home visits

Wolf et al. (2019) has 5 study arms

Arm 1: None

Arm 2: Teacher training and coaching program + Reinforcement text messages to teachers

Arm 3: Teacher training and coaching program

Arm 4: Teacher training and coaching program + Parent awareness meetings + Reinforcement text messages to teachers + Reinforcement flyers to parents

Arm 5: Teacher training and coaching program + Parent awareness meetings

VI. Number of rounds of data collection

- **Definition:** Number of rounds of data collection including baseline in the study
  - Round: Collects data from the same **data source** at a **given time**
    - What we want to know: when are different outcomes for which there is a treatment effect estimated measured?
      - Data source: surveys vs. administrative data (usually)
      - Samples: different survey samples conducted at same time should be considered the same round
      - Outcomes: only outcomes for which there are treatment effects estimated in the main paper

Example

- **Blattman et al. (2020) have 4 rounds of data collection**

We randomly sampled 5 people per group (2,677 people) for a baseline survey and attempted to track this sample over time. We conducted the baseline post-randomization due to government funding delays, and three groups (3 percent of the sample) could not be surveyed.

The government disbursed grants between July and September 2008. The two-year end line was conducted between August 2010 and March 2011, 24 to 30 months after disbursement; a four-year survey was conducted between April and June 2012, 44 to 47 months after disbursement; a nine-year end line was conducted between March and May 2017, 103 to 106 months after disbursement.

Example

Chong et al. (2015) has 2 rounds of data collection

1. Election data
2. Follow-up survey

Not census data, candidate data

We collected four types of data. First, electoral results at the precinct level come from state electoral institutes, which also provided maps and geo-referenced voting precincts that we used for the distribution of the flyers. Second, demographic baseline characteristics come from census data originally reported at the village and block levels. Villages

the villages (or blocks) inside the voting precinct. Third, we conducted a survey approximately two weeks after the interventions (10 days after the elections). Since our intent

Finally, we collected data on candidates' jobs prior to the 2009 election.

Example

Muralidharan et al. (2021) has 7 rounds of data collection

1 Register of landlords

2, 3, and 4: Bank-based dataset July, August, and September 2018

5, 6, and 7: MAO record dataset July, August, and September 2018

Not phone calls or survey

We primarily measure outcomes using administrative data, including (i) the register of all agricultural landholders in the state, including names, village, acres held, and a contact phone number; (ii) a farmer-level record of check distribution maintained by the MAOs;[22] and (iii) farmer-level bank records of check encashment. Our analysis focuses on encashment, as getting the money is the ultimate

We received the up-to-date MAO and bank-based databases at three points in time: once in July, once in August, and once in September 2018. When we compare the

X We also use data from phone calls conducted by the call center as a secondary data source. These data were collected over the phone from 22,127 program

X Finally, we use data from a short phone survey of MAOs.

Poll

Wolf et al. (2019): How many rounds of data collection?

A. 2
B. 3
C. 4
D. 5

Measures

All measures were collected at baseline (beginning of the school year) and follow-up (end of the school year). All data within a school—teacher surveys, classroom observations, and child outcomes—were collected during the same visit at each round.

Poll

Asunka et al. (2019) How many rounds of data collection?

A. 1

B. 2

C. 3

D. 4

Our first indicator of fraud is polling station *turnout*. [75] If turnout correlates with the placement of observers, this is evidence of fraud because it shows that turnout is artificially high at stations without observers. To construct this measure, we obtain data on the number of votes cast in each of our sample polling places using information from the official polling station results forms. Turnout is calculated as the total number of valid and rejected ballots divided by

To measure election violence, we code whether a polling station experienced *intimidation during voting*. This outcome is one indicator from a survey that we conducted at each station in the sample. At treated stations, the information we use was reported by CODEO observers as

FYI for Stage 3

- We will ask additional information about data collection timing in Stage 3
  - **Data collection start date**
    - Blattman et al. 2-year endline: August 2010
  - **Data collection end date**
    - Blattman et al. 2-year endline: March 2011
  - **Data collection end date calculated from duration**
    - No (end date given by authors)
  - **Note:** authors do not always include this information, especially for administrative data
    - -99 is an option for day, month, and year

Thank you for listening

Jennie Barker
jlbarker@berkeley.edu

Impact Data and Evidence Aggregation Library