Sampling

Alaka Holla

June 12, 2025

GOAL

We want to know who do the average treatment effects pertain to.

How much can we generalize effects?

One way to understand this is to understand who exactly is in the experimental sample.

Sampling fields under construction

1. Testing simplest way to do this

2. First get familiarity describing *randomized assignment to arms* (which uses the same terms to describe a different process)

Agenda

1. Representativeness
2. Ways of selecting units
3. Stratification in sampling and Inclusion/exclusion criteria
4. Proportions and counts
5. Examples

1 **Representative samples**

A representative sample

A subset of a population that accurately represents or reflects the characteristics of the population.

Typically needed:
Random sampling
Proportionate sampling

Representative of..

- Country
- Women ages 15-49
- Public schools
- Female entrepreneurs without access to credit
- NGO clients

A representative sample

Experimental samples are rarely representative of a very broad population. They are *convenience* samples.

But they may be representative of a very narrowly defined population.

Goal: Precisely describe the population that is in the sample. How did they get there?

Constraints

- Need to work with a certain implementer or in certain places
- Consent
- Safety
- Survey logistics
- Interest in subpopulations

Whom does this generalize to?

Women of childbearing age in a regionally representative household survey

OR

Recent or expectant mothers attending prenatal care in 10 public clinics

American Economic Journal: Economic Policy 2025, 17(2): 388–417
https://doi.org/10.1257/pol.20230283

Informing Mothers about the Benefits of Conversing with Infants: Experimental Evidence from Ghana†

By Pascaline Dupas, Camille Falezan, Seema Jayachandran, and Mark Walsh*

We evaluate a low-cost intervention designed to boost parents' verbal engagement with infants, which tends to be limited in developing countries. In our randomized experiment, recent or expectant mothers watched a three-minute informational video and received a themed calendar. Six months later, treated mothers reported stronger belief in the benefits of verbal engagement, more frequent parent-infant conversation, and more advanced infant language skills. Treatment effects on objective measures of parent-child conversation frequency and infant skills were positive but insignificant. We find larger immediate treatment effects on objective parent-child conversation, suggesting potentially larger long-term effects had the behavior change stuck more. (*JEL* D83, D91, I26, J13, J16, O12)

Whom does this generalize to?

To answer this, we do not need a geographically representative sample.

II. Study Design

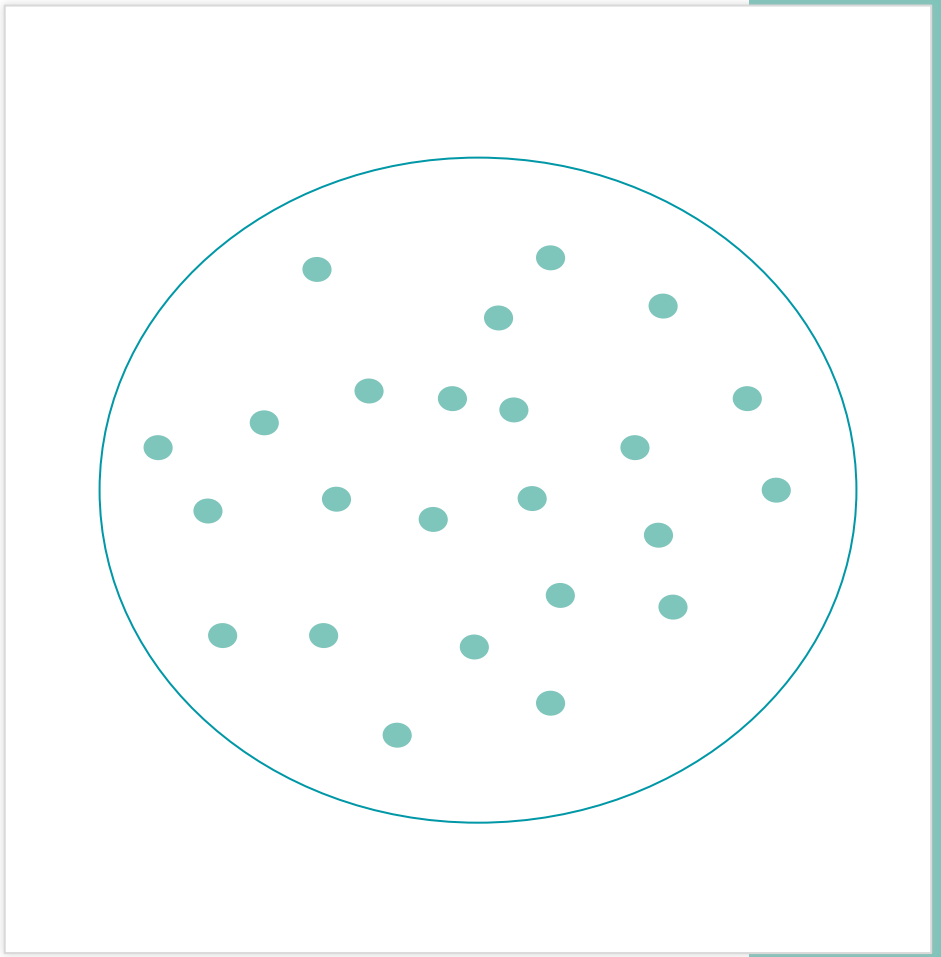
A. *Sampling and Intervention*

We received approval from the Ghana Health Service, the government agency overseeing health clinics, to survey prenatal and postnatal patients in ten of the public health clinics around the city of Tamale in early 2021 (see Supplemental Appendix Table A.1 for the list of facilities). Tamale is the third-largest city in Ghana and the largest city in the Northern Region of Ghana, which is poorer than the rest of Ghana.[5]

In March 2021, we employed a team of surveyors from Innovations for Poverty Action (IPA) Ghana to enroll a sample of prenatal and postnatal patients from the health clinics. Surveyors approached patients before/after their prenatal or postnatal clinic visits and, if the patients were interested, screened them for eligibility. In order to participate, women had to (i) be aged 18–40 years old, (ii) have an infant or be pregnant with a child who would be 2–18 months old at the time of the follow-up survey 6 months later, and (iii) speak English or Dagbani (the main language in Tamale).[6] We aimed to survey 1,400 women and ended up surveying slightly more: 1,408.

2 Ways of selecting units

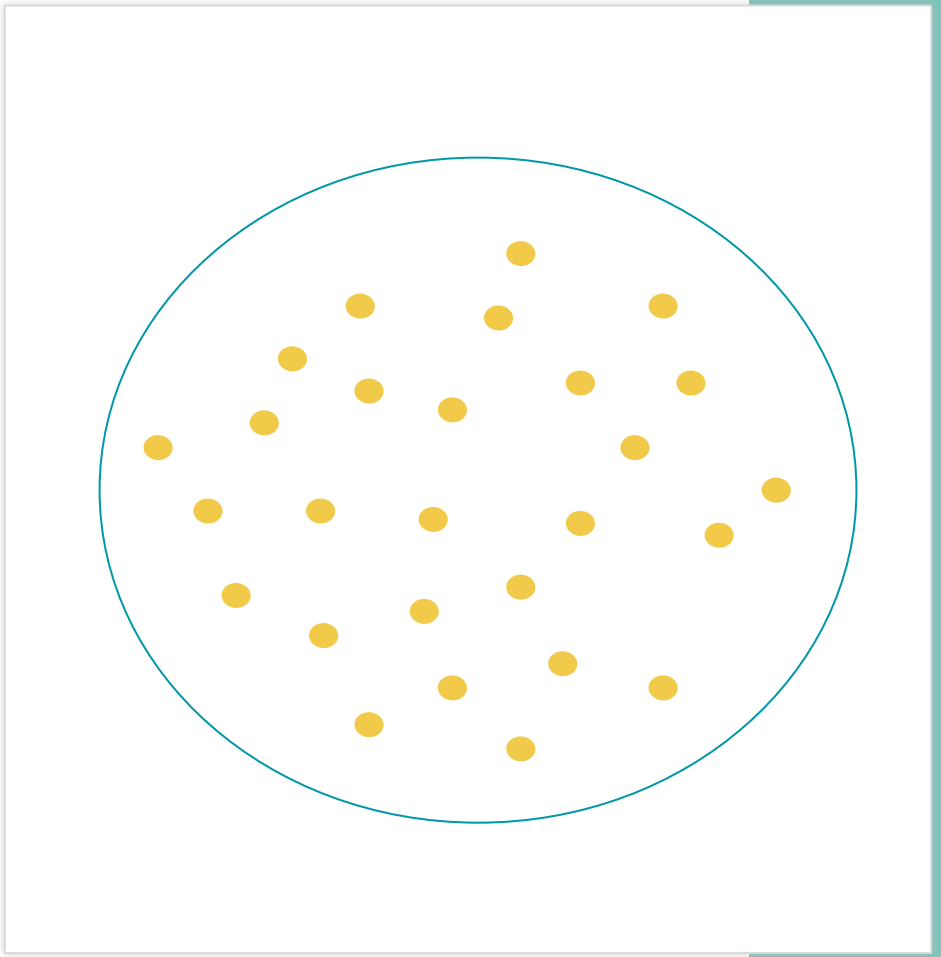
How are units chosen?



How are units chosen?

Universe

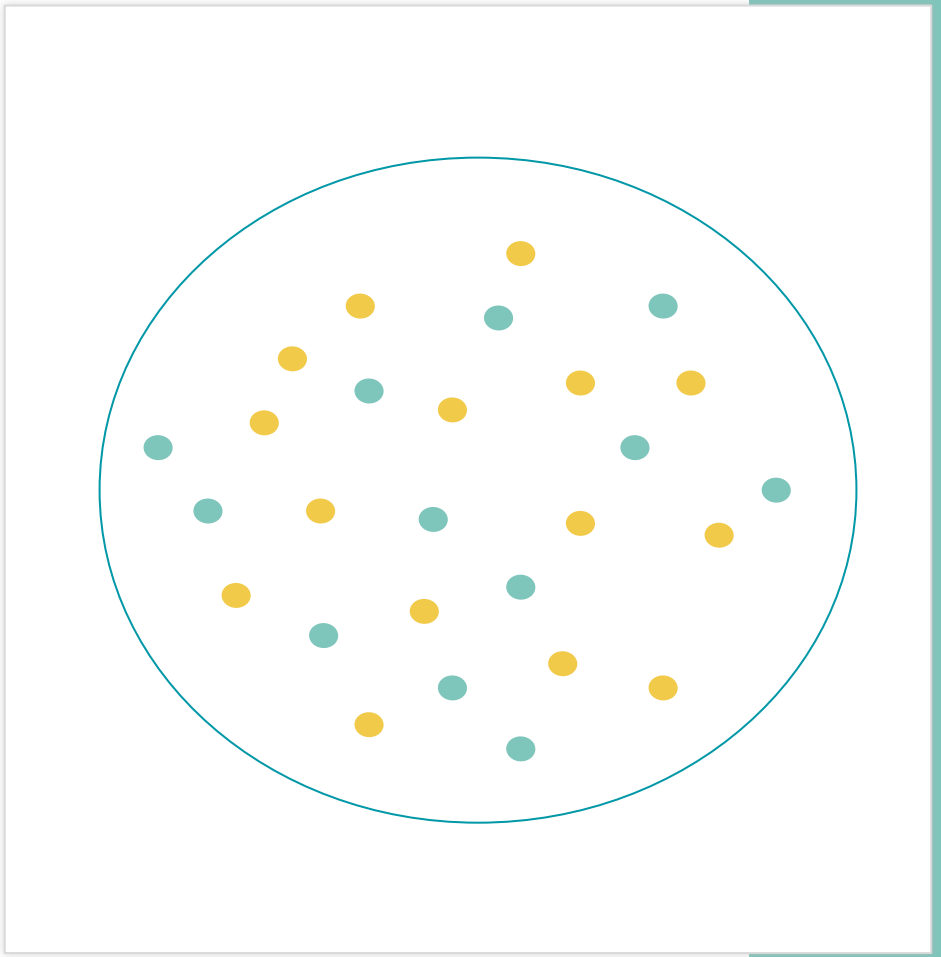
Select all units



How are units chosen?

Random

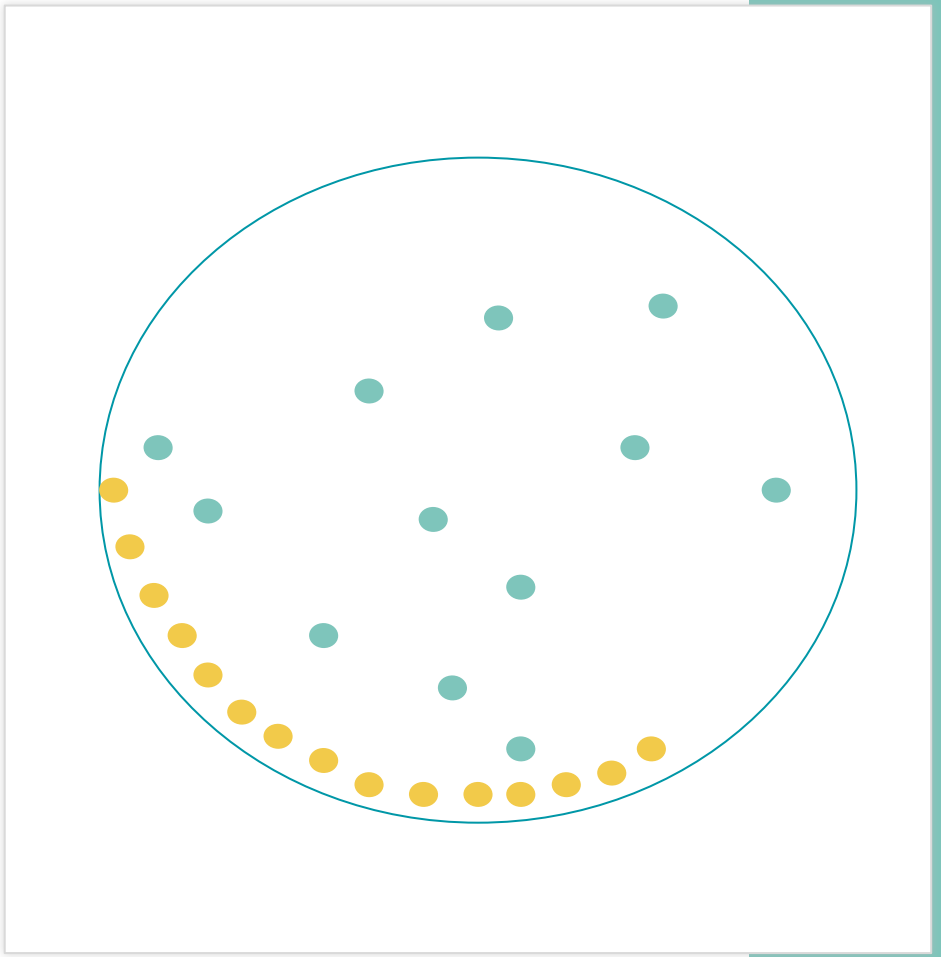
Randomly select units



How are units chosen?

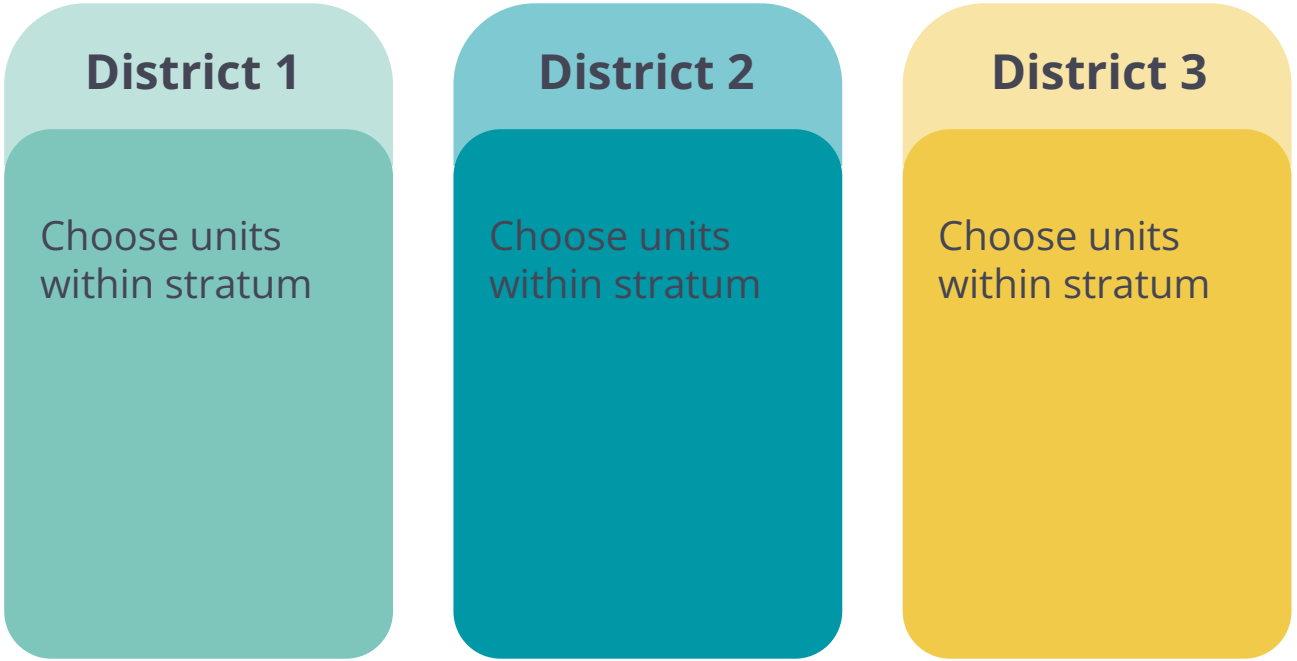
Non-random

All the people who live next to the road



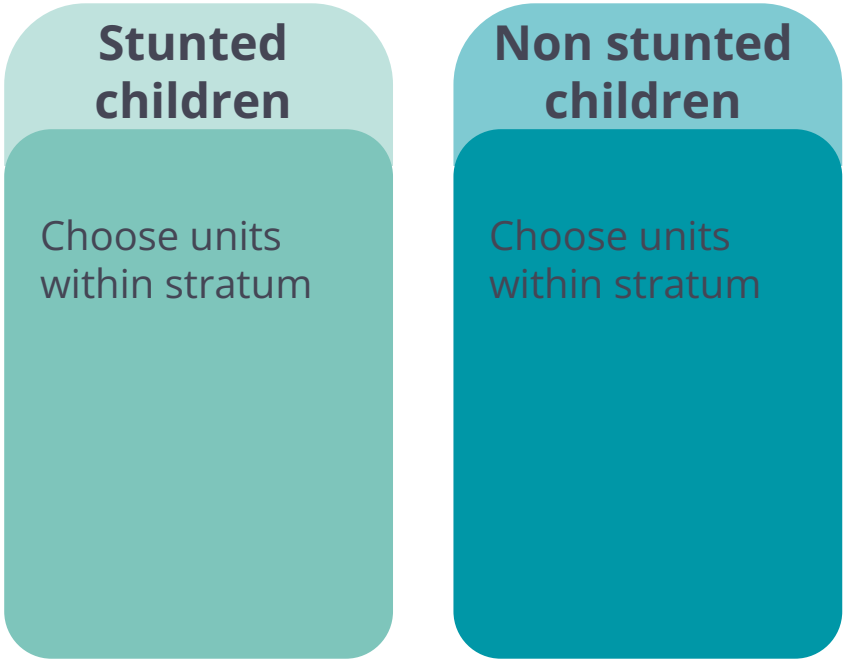
3 Stratification and inclusion/exclusion

Stratification - *Example: Geography*



Why
Ensure sample from all 3 districts.

Stratification - *Example: Nutritional status*



Inclusion

Selected units have to have this trait.

Examples

Clinics have to have staff

Household has to have a child under age 5

Women have to be pregnant

Exclusion

Selected units cannot have this trait.

Examples

Population size under 100

Child with a known developmental delay

Centers with a similar program

4 Proportions or counts

Proportion

Number of units selected varies based on size of larger unit

Examples

Choosing more households in larger villages than in smaller villages. Exact number based on village's share in total population.

Count

Choose a fixed number of units to select

Examples

Choosing 40 households per village.

Why do we want to know all this?

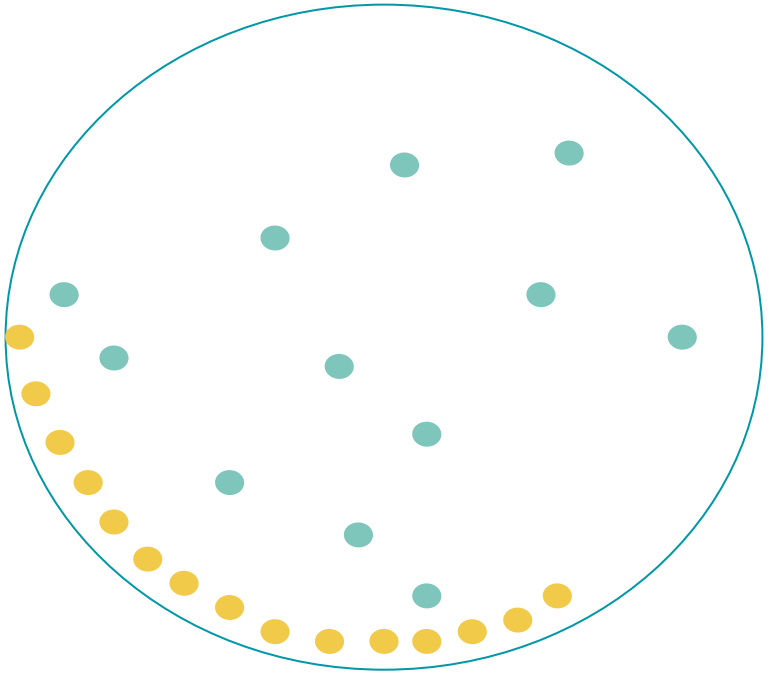
Say there are 3 districts in a province.

If districts do not have same population, then choosing same # of units per district → population not representative of province.

Why do we want to know this?

Who are results relevant for:
All the people who live next to the road

NOT: All the people in the larger unit



5 Examples

Warning!

Papers do not always present this information:

1. At all
2. Clearly
3. Using these terms in this sequence

Who do these results pertain to?

- How are units selected?
- Was there any stratification *for sampling*?
- Were there any inclusion or exclusion criteria?
- Did they sample in proportion or using a fixed number?

Augmenting State Capacity for Child Development: Experimental Evidence from India

Alejandro J. Ganimian

New York University

Karthik Muralidharan

University of California, San Diego

Christopher R. Walters

University of California, Berkeley

We use a large-scale randomized experiment to study the impact of augmenting staffing in the world's largest public early-childhood program: India's Integrated Child Development Services. Adding a worker doubled net preschool instructional time and led to increases of 0.28σ and 0.46σ in math and language test scores after 18 months for children who remained enrolled in the program and 0.13σ and 0.10σ for all children enrolled at baseline. Rates of stunting and severe malnutrition were also lower in the treatment group for children who remained enrolled. A cost-benefit analysis suggests that the benefits of augmenting staffing significantly exceed its costs. These effects are likely to replicate even at larger scales of program implementation.

What is being selected?

Unit of randomization

Anganwadi center

Unit of analysis

Anganwadi worker, child

Augmenting State Capacity for Child Development: Experimental Evidence from India

Alejandro J. Ganimian

New York University

Karthik Muralidharan

University of California, San Diego

Christopher R. Walters

University of California, Berkeley

We use a large-scale randomized experiment to study the impact of augmenting staffing in the world's largest public early-childhood program: India's Integrated Child Development Services. Adding a worker doubled net preschool instructional time and led to increases of 0.28σ and 0.46σ in math and language test scores after 18 months for children who remained enrolled in the program and 0.13σ and 0.10σ for all children enrolled at baseline. Rates of stunting and severe malnutrition were also lower in the treatment group for children who remained enrolled. A cost-benefit analysis suggests that the benefits of augmenting staffing significantly exceed its costs. These effects are likely to replicate even at larger scales of program implementation.

What do we see here?

There is random selection

BUT

Stratification (district, project, vacancy)

Exclusion restrictions

Not so relevant representativeness claim

A. Sampling, Randomization, and Implementation Quality

We randomly sampled four districts across the state to be representative of a population of 60 million people.[9] For ICDS administration, each district is divided into projects comprising 100–150 *anganwadi* centers, which are

[8] See https://www.socialscienceregistry.org/trials/1772.

[9] We excluded the district of Chennai, which is the state capital and a metropolis of more than 7 million people. District sampling was stratified by geographic zones and average nutrition status (see fig. A.1, in app. A; apps. A–D, including figs. A.1–C.1, are available online). Table B.1, in app. B (tables A.1–C.1 are available online), shows that the four sampled districts are very similar to nonsampled districts. For inference, we condition on the

We started with the universe of AWCs in the four sampled districts and excluded those with other NGO interventions, in buildings shared with other centers, and with vacancies in both staff positions (worker and helper).[10] We then randomly sampled 320 centers from the remaining population, stratifying by staffing vacancy and project.

How many stages of selection were there?

District - random, after stratification and exclusion, fixed #

Centers - random, after stratification and exclusion, fixed #

Children 1 - universe
Children 2 - random, fixed proportion

Who do these results pertain to?

- How are units selected?
- Was there any stratification *for sampling*?
- Were there any inclusion or exclusion criteria?
- Did they sample in proportion or using a fixed number?

American Economic Review 2012, 102(4): 1206–1240
http://dx.doi.org/10.1257/aer.102.4.1206

Targeting the Poor: Evidence from a Field Experiment in Indonesia†

By Vivi Alatas, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias*

This paper reports an experiment in 640 Indonesian villages on three approaches to target the poor: proxy means tests (PMT), where assets are used to predict consumption; community targeting, where villagers rank everyone from richest to poorest; and a hybrid. Defining poverty based on PPP$2 per capita consumption, community targeting and the hybrid perform somewhat worse in identifying the poor than PMT, though not by enough to significantly affect poverty outcomes for a typical program. Elite capture does not explain these results. Instead, communities appear to apply a different concept of poverty. Consistent with this finding, community targeting results in higher satisfaction. (*JEL* C93, I32, I38, O12, O15, O18, R23)

What is being selected?

Unit of randomization

Subvillage

Units of analysis

Subvillage, village head, households

American Economic Review 2012, 102(4): 1206–1240
http://dx.doi.org/10.1257/aer.102.4.1206

Targeting the Poor: Evidence from a Field Experiment in Indonesia†

By Vivi Alatas, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias*

This paper reports an experiment in 640 Indonesian villages on three approaches to target the poor: proxy means tests (PMT), where assets are used to predict consumption; community targeting, where villagers rank everyone from richest to poorest; and a hybrid. Defining poverty based on PPP$2 per capita consumption, community targeting and the hybrid perform somewhat worse in identifying the poor than PMT, though not by enough to significantly affect poverty outcomes for a typical program. Elite capture does not explain these results. Instead, communities appear to apply a different concept of poverty. Consistent with this finding, community targeting results in higher satisfaction. (*JEL* C93, I32, I38, O12, O15, O18, R23)

What do we see here?

Unit of randomization: sub-village

Province: non-random

Village: random with stratification and exclusion

Sub-village: random

B. *Sample*

The sample for the experiment consists of 640 subvillages spread across three Indonesian provinces: North Sumatra, South Sulawesi, and Central Java. The provinces were chosen to represent a broad spectrum of Indonesia's diverse geography and ethnic makeup. Within these three provinces, we randomly selected a total of 640 villages, stratifying the sample to consist of approximately 30 percent urban and 70 percent rural locations.[8] For each village, we obtained a list of the smallest administrative unit within it (a *dusun* in North Sumatra and *Rukun Tetangga* (RT) in South Sulawesi and Central Java), and randomly selected one of these subvillages for the experiment. These subvillage units are best thought of as neighborhoods. Each subvillage contains an average of 54 households and has an elected or appointed administrative head, whom we refer to as the subvillage head.

[8] An additional constraint was applied to the district of Serdang Bedagai because it had particularly large-sized subvillages. All villages in this district with average populations above 100 households per subvillage were excluded. In addition, five of the originally selected villages were replaced prior to the randomization due to an inability to reach households during the baseline survey, the village head's refusal to participate, or conflict.

What do we see here?

Unit of analysis:

Household

Random (+ inclusion?)

Fixed count

A. *Data Collection*

We collected four main sources of data: a baseline household survey, household rankings generated by the treatments, data on the community meeting process (in community/hybrid treatments only), and data on community satisfaction.

Baseline Data.—We conducted a baseline survey in November and December 2008. The survey was administered by SurveyMeter, an independent survey organization. At this point, there was no mention of the experiment to households.[17] We began by constructing a complete list of all households in the subvillage. From this census, we randomly sampled 8 households from each subvillage plus the head of the subvillage, for a total sample size of 5,756 households. To ensure gender balance among survey respondents, in each subvillage, households were randomized as to whether the household head or spouse of the household head would be targeted as the primary respondent. The survey included questions on demographics, family networks in the subvillage, participation in community activities, relationships with local leaders, access to existing social transfer programs, and households' per capita consumption.

[17] SurveyMeter enumerators were not told about the targeting experiment.

Papers are not super clear.

We will document what is mentioned in the study, even if it's in footnote 93.

What’s coming next

Sampling survey fields (post asynchronous training)

Training and practice on sampling fields

Integration of sampling fields into full IDEAL survey

Thank you for listening

Alaka Holla
aholla@worldbank.org

IDEAL
Impact Data and Evidence Aggregation Library