Impact Evaluation in Education

Focusing on managing Rigorous Evaluation (RE)

Introduction to RCT and Research Design

By

Minhaj Mahmud

IGS-BDI, Brac University

March 23-24, 2014
BRAC Inn Centre, Dhaka, Bangladesh.

Randomized Controlled Trial(RCT)

- **RCT is regarded as the most powerful research design in drawing conclusions about the impact of any intervention on specific outcome**
  - **To understand whether public programs actually work as they are designed to reach certain goals and beneficiaries**
- **Implementation and evaluation by comparing different treatment groups chosen at random of an intervention or a set of intervention specifically designed to test a hypothesis or a set of hypothesis"**
  - **"such design also facilities ethical allocation decision when facing resource and or time constraints."**

Why Impact Evaluation?

- **Central issue in policy design; Policy questions are causal in nature**
- **Questions we can answer with randomized evaluations?**
  - **(How) does incentive work in developing countries institutions?**
  - **How much does an education program improve test score?**
  - **Conditional cash transfer program improve health and education outcome ?**
  - **The question of behavior change: why are not people doing things that are obviously good for them?**

How is Impact Measured?

- **How much does an education program improve test score?**
- **What is the test score with and without program exposure?**
  - **Compare the same individual with and without programe at same point of time**
  - **Can we observe same individual with and without program at same point in time?**
  - **Need counterfactual,(key thing) (Example)**
  - **So the estimated impact is difference between treated and counterfactual**
- **Formally impact is : $\alpha = (Y \mid P=1) - (Y \mid P=0)$**

What is commonly done?

- **Before and after comparisons(compare outcome of interest before(t=0) and after (t=1)**
  - **Impact of intervention plus whatever happened between t=0 and t=1**
  - **Common time effects: Underestimate or overestimate effects , magnitude and sign of effect**
- **Participants and non-participant comparisons**
  - **Selection bias: Intervention effects plus whatever is different between participants and non-participants**
  - **Make it impossible to disentangle unobservable characteristics of individuals from intervention impacts**
  - **Overestimate and underestimate impacts**
- **Statistical correlation**
  - **Multiple regression context cannot handle – induces correlation between intervention and regression errors**

Randomized Trials

- **Evidence about counterfactuals is often generated by randomized trials or experiments**
- **Eliminates common biases (or confounders) when done properly**
  - **Selection bias**
  - **Trends concurrent with intervention**
- **Therefore, RCT is often considered the gold standard of estimating causal impacts**

Impact of P on Y?

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

OBSERVE $(Y \mid P=1)$
Outcome with treatment

ESTIMATE $(Y \mid P=0)$
The Counterfactual

IMPACT = Outcome with treatment − counterfactual

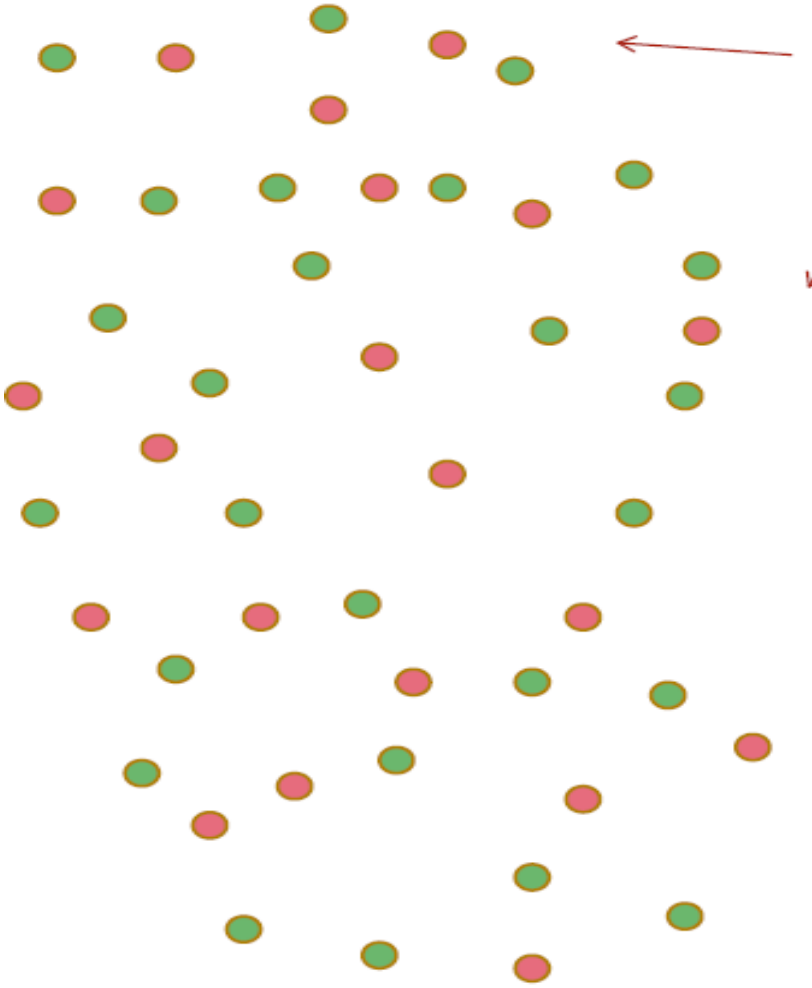
- Intention to Treat (**ITT**) – *Those offered treatment*
- Treatment on the Treated (**TOT**) – *Those receiving treatment*

- Use **comparison** or **control** group

What does randomization achieve?

- **Overcome potential confounders**
- **Selection bias**
  - **Participants might be characteristically different from non- participants**
  - **Give all equal chance of being in control or treatment groups**
  - **Guarantees that all other factors will be on average equal between groups**
- **Only difference is the intervention**
  - **Treatment and control group differs in expectation only through exposure to treatment condition.**
  - **So outcome is same for both in the absence of treatment**

Selection bias



- Green = treatment (with intervention)
- Pink = comparison (without intervention

- Should average characteristics differ across treatment and comparison groups prior to the intervention?
  - No.

Randomization solves selection bias

- **Differences in average untreated outcome between treatment and control group**
  - **Average characteristics should be the same for treatment and comparison groups prior to the intervention**
  - **For example prior to a health insurance intervention, average expenditure ($\bar{e}$) should be identical in treatment and comparison groups**
  - **Common trends/factors -When treatment units selected randomly, shocks are common to both treatment and control group so that difference between groups at the end is due to intervention**

Randomization solves selection bias

$$\begin{aligned}\overline{\delta} &= E_U[Y_1(u)\,|\,D=1] - E_U[Y_0(u)\,|\,D=0] \\ &= E_U[Y_1(u)\,|\,D=1] - E_U[Y_0(u)\,|\,D=0] + E_U[Y_0(u)\,|\,D=1] - E_U[Y_0(u)\,|\,D=1] \\ &= E_U[Y_1(u) - Y_0(u)\,|\,D=1] + E_U[Y_0(u)\,|\,D=1] - E_U[Y_0(u)\,|\,D=0]\end{aligned}$$

- **Since selection bias is zero we get an unbiased estimate of treatment impact**
- **Control variables should not affect bias (?) and increase the precision of estimate**

Unit of Randomization

- **Extremely important because it determines**
  - **The extent randomization can solve selection bias**
  - **Statistical power**
  - **Ability to measure externality**
- **Consider one treatment and one control unit. What happens if one face some shocks ?**
  - **Can we disentangle treatment effects from shock effect? Treatment and control units are unlikely to be balances on average characteristics**
  - **What if T=5 and C=5?**

Unit of Randomization

- **As a rule of thumb, randomize at smallest feasible units of implementation**
  - **Choose according the type of program**
  - **Sufficiently large number of unit**
- **Remember to detect desired impact(power), spillover, and above all cost of implementation**

Statistical Power

- **The power is the probability that we can detect an impact, when in fact one exists.**
- **The null hypothesis is that the program does not have an impact(Ho: impact = 0)**
- **The alternative hypothesis is that the program has an impact(Ha: impact ≠ 0)**
  - **N should be such that we won't mistakenly find an impact when there is not actually one**
  - **We are able to find an impact when there is one**
  - **The risk of Type II errors can be reduced if the impact evaluation has a high *power.***
- **Identify the sample size to detect whether a given change in the outcome of interest is statistically significant.**
- **Identify the needed size of the change in outcome to detect its statistical significance given a certain sample size.**

What do we need for power calculation?

What ?	How ?
Significance level	The lower it is, the larger the sample size needed for a give power (usually at 5%)
The mean and the variance of the outcome in the comparison group	The larger the variability is, the larger the sample for a given power(based on previous surveys in similar settings)
The size of the impact that we want to detect	The smaller the impact we want to detect, the larger a sample size we need for a given power (set to the smallest impact)
Statistical power calculations can take into account clustering(intra-cluster correlation) in estimating the number of clusters and number of observations.	

Statistical Power

- **What is the risk of underpowered study?**
  - **Does not have enough observations to detect an impact of a theorized size), then probably best to rethink whether to pursue the evaluation.**
  - **If data collection is involved, and we have an underpowered study, it could turn out to be a futile and costly (time, labor, etc.) to do it.**
- **Remember: In RCTs: The effective sample size is at the unit of randomization.**
  -

Note on statistical power

- **How to increasing power of study?**
  - **Note that power calculations is a rule of thumb.**
  - **Control for other relevant factors (X's) in your treatment-effect regressions.**
  - **Stratification, paired randomization and inclusion of such data in treatment-effect regressions.**
  - **Limit survey attrition.**
- **Although a technical issue, power calculations carry large practical benefits**

Stratification

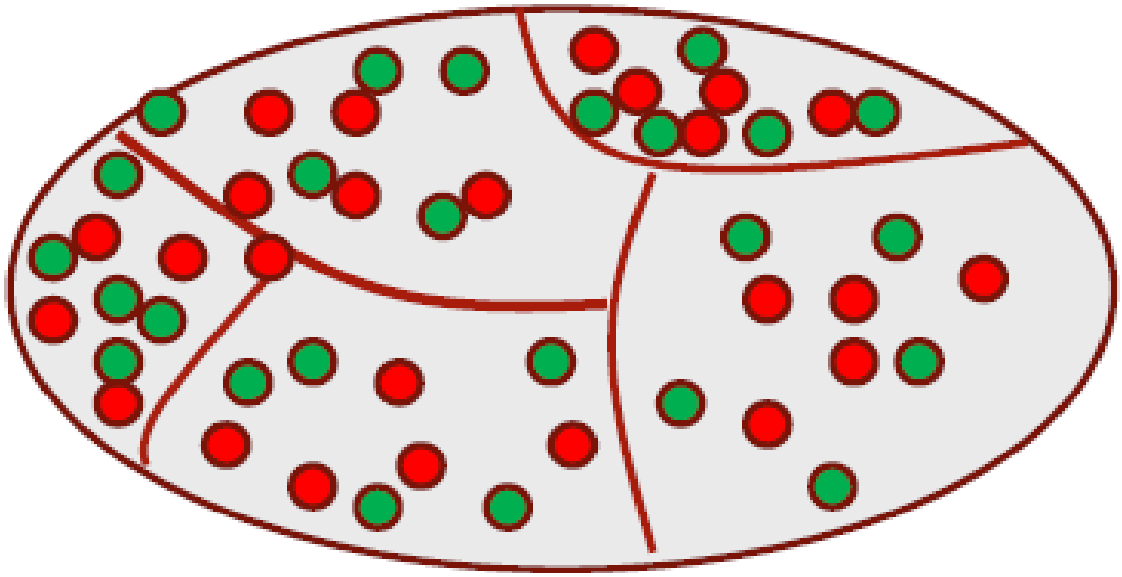
- **What is the impact in particular region?**
  - **Random assignment to treatment within geographical units(other subpopulation)**
  - **Equally distribute to treatment and controls**
- **Measure heterogeneous treatment effects**
  - **balance stratified variables between treatment and controls and improves power**

Stratification and Randomization

- **Separate units into sub-population**
  - **e.g. Geographic, Gender, Ethnicity, Income**
- **Randomize treatment with each strata**

■ ● = T

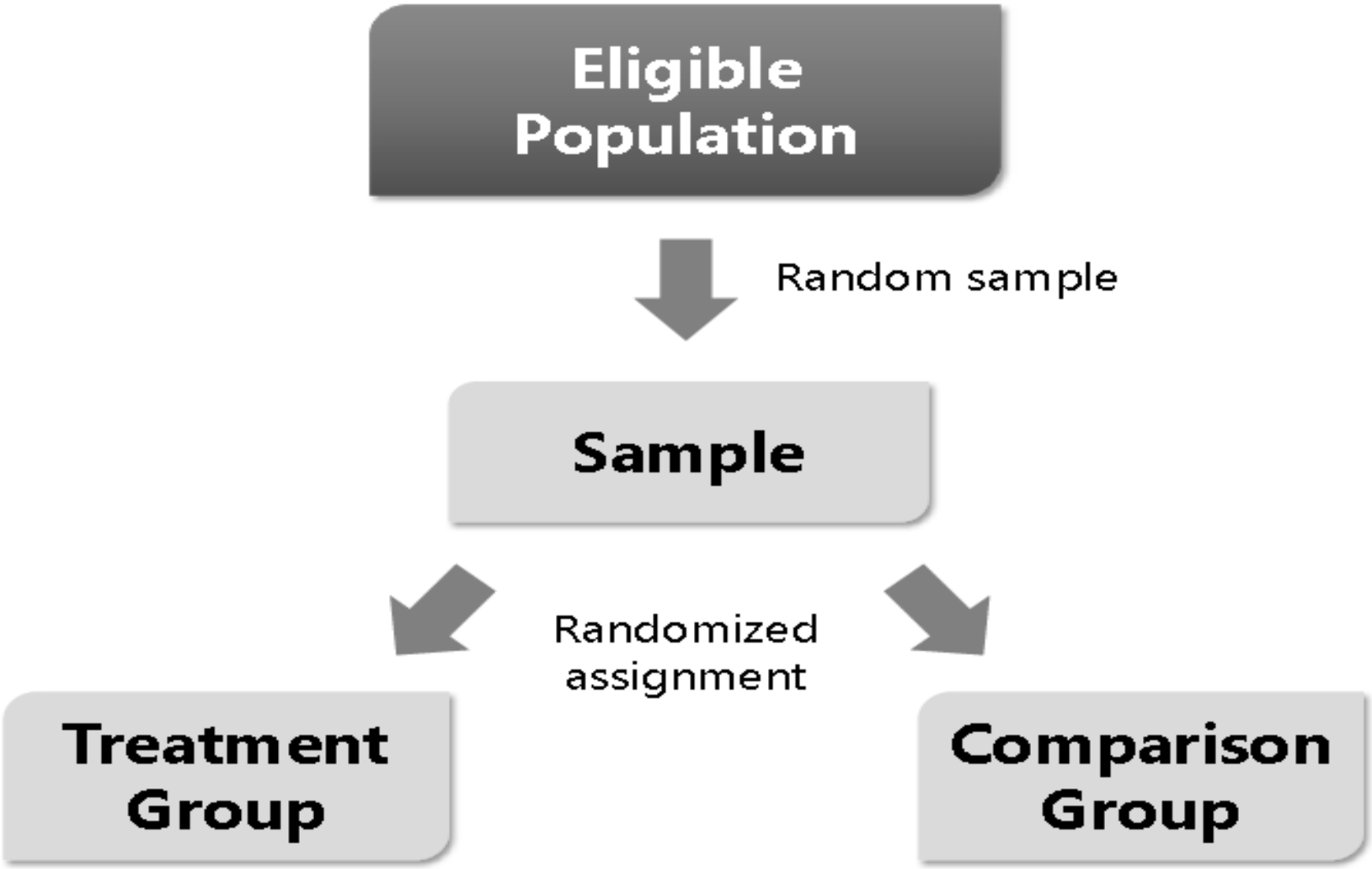
■ ● = C



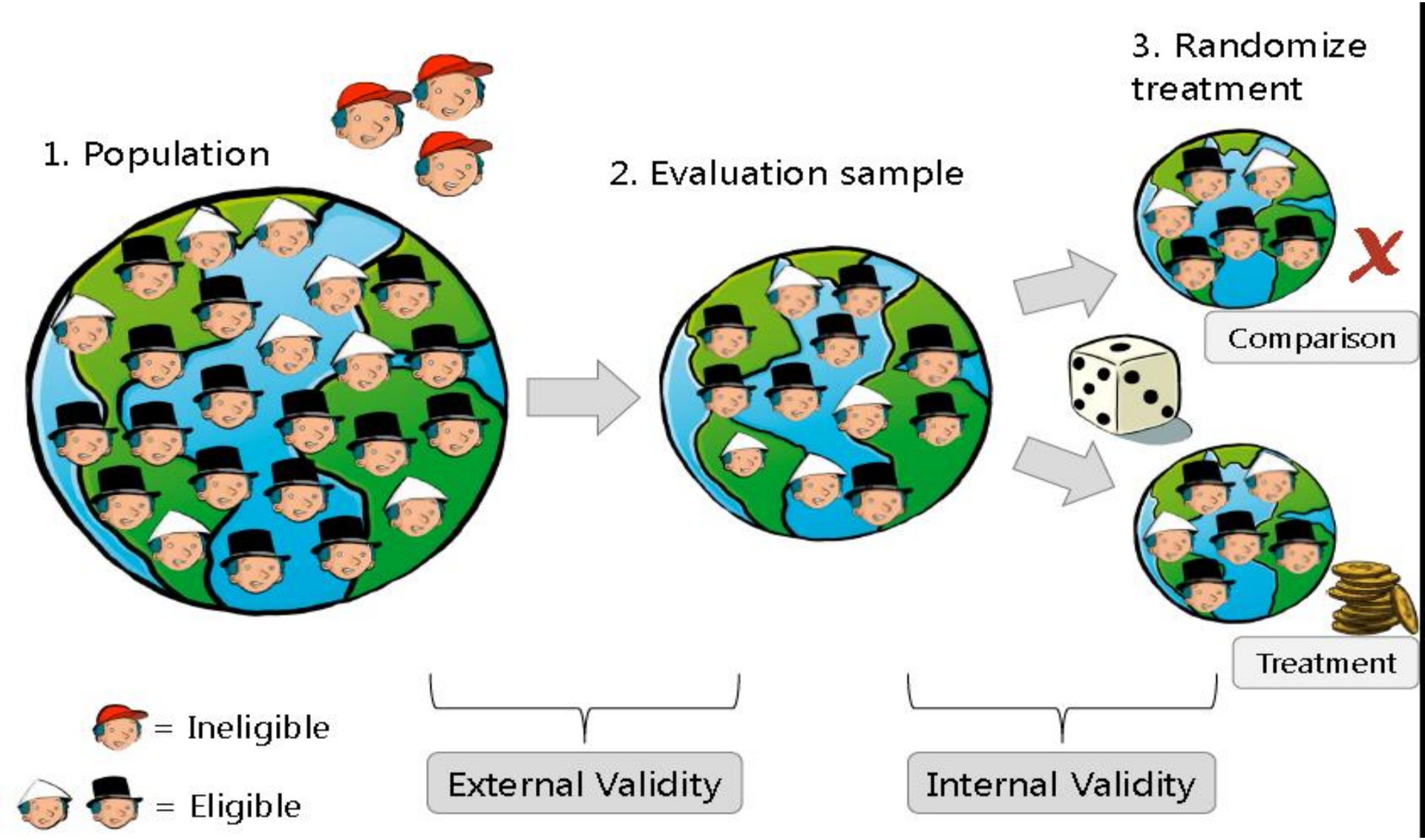
Random Sampling

- **Randomization : Random assignment to treatment and control group**
  - **Unbiased treatment of impact for the sample in question**
- **Randomly choosing units from overall population**
  - **Simple lottery:**
  - **Consider all in the population of interest and place all names written in piece of papers in a box- then draw half of the names-to offer intervention**
- **Sampling before or after assignment of treatment condition**
  - **in case of large intervention after assignment**
  - **Do not need to survey everyone to estimate treatment effect**

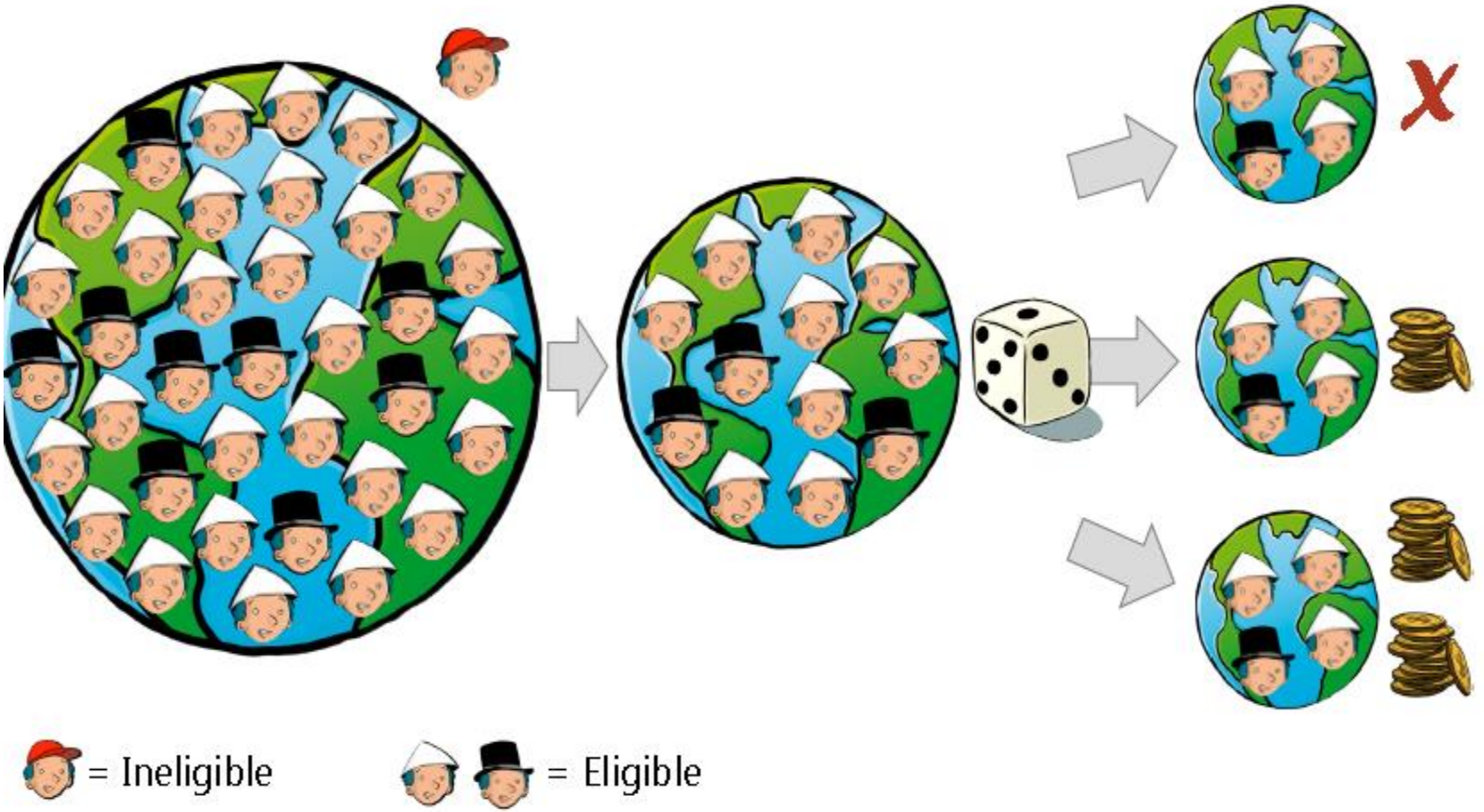
Sampling and Randomization



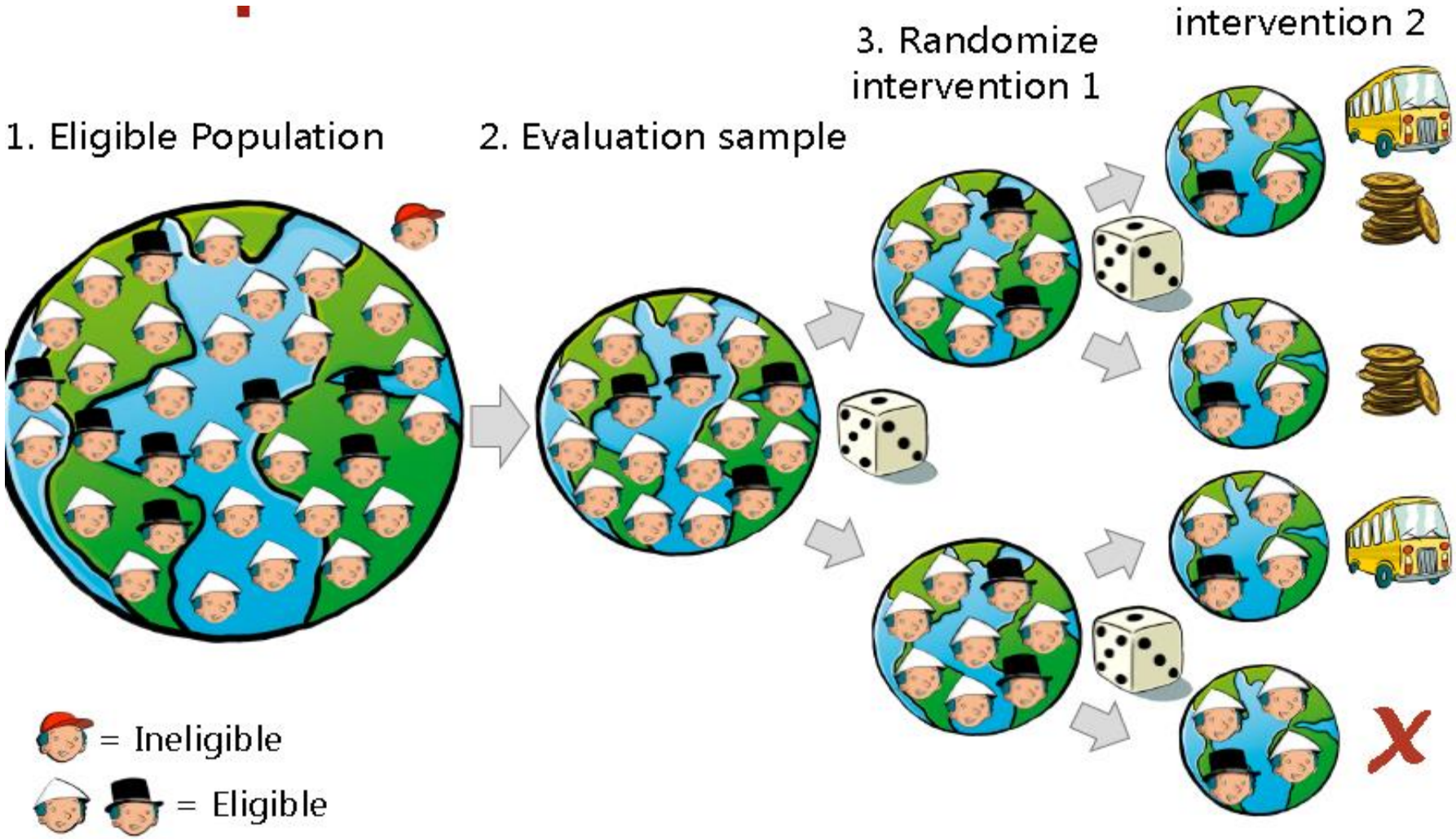
How does Randomized Trial Look?



Randomized trial: impact of different levels of treatment



Randomized trial: Impact of an intervention compared the other



Issues/Concern in RCT

- **Attrition**
  - **Drop out from treatment or sample**
  - **Overestimate treatment effects , if only higher beneficiaries remain**
  - **Underestimate treatment effects if a group had to migrate out from control**
- **Matter of concern**
  - **Differential attrition across treatment and control group**
  - **Differential attrition across types within a experimental group**

Issues/Concern in RCT

- **Non-compliance**
  - **Some treatment member opts out treatment**
  - **Some control get access to treatment**
- **Results within experimental unit or implementer**
- **Not random- may affect causal inference**
- **Often unavoidable but smaller occurrences can be addressed**

Issues/Concern in RCT

- **Wrong parameter estimates**
  - **Randomization can solve selection bias but it may be still possible that we are estimating something different then we intended for**
- **Partial or total derivative?**
  - **Direct plus indirect impact versus direct impact?(Example?)**
- **Evaluation design should take this into account**

Summing up

- **Randomization solves selection bias**
- **Randomization cannot solve all problems**
  - **Statistical power**
  - **Attrition and no-compliance**
  - **Potential deviation from parameters of interest**
- **Need large sample size**
- **May not be feasible (politically)**

Ethical Considerations

- **Do not delay benefits**
  - **Rollout based on budget/administrative constraints**
- **Equity**
  - **Equally deserving beneficiaries deserve an equal chance of going first**
  - **Give everyone eligible an equal chance**
- **Transparent**
  - **If rank based on some criteria, then criteria should be quantitative and public**

Impact of parental training on Early childhood stimulation

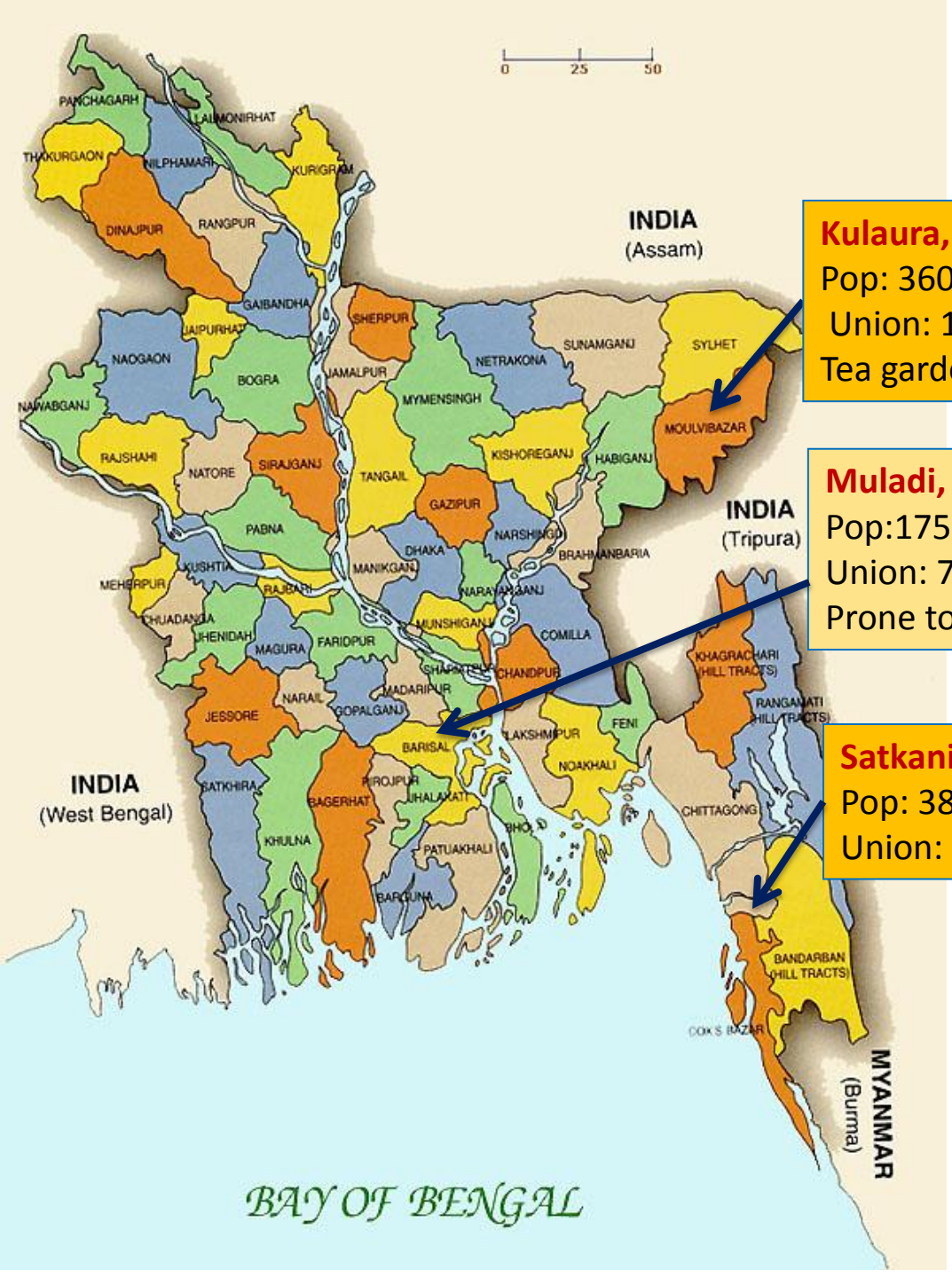
General Objective:

Determine the effectiveness of an innovative program that provides education to families about early childhood stimulation as an add-on to a national early childhood nutrition program.

Implementing agency: Save the Children

Evaluation agency: American Institute for Research (AIR)

Program Areas



Kulaura, Maulvibazar
Pop: 360000
Union: 13
Tea gardens: 15

Muladi, Barisal
Pop:175000
Union: 7
Prone to river-erosion

Satkania, Chittagong
Pop: 385000
Union: 17

Selection Criteria:

1. Not currently implementing NNS
2. Low presence of other actors
3. Geographically distinct to maximize use of learning nationally

National Nutrition Service (NNS) and Early Stimulation Package

National Nutrition Services

Community level

- Screening for malnutrition
- Infant Young Child Feeding counseling
- Specific breastfeeding counseling
- Micro-nutrient supplementation
- Referral
- Follow-up home visits

Household Level

- Screening for malnutrition
- Referral
- Treatment of malnutrition at home
- Nutrition advice
- BCC activities

Early Stimulation Services

Key Messages dissemination and services for Mothers and Caregivers on:

- Care during pregnancy (for pregnant women)
- Love and affection to the child
- Play and games
- Talking and communicating
- Positive discipline
- Responsive feeding
- Health and hygiene
- Share messages with others

(using Child Development Card, Key Message Booklet, Children picture books)

Research Questions

- What is impact of the early childhood stimulation program on children's *cognitive development outcomes*, anthropometric *outcomes*, mothers' *parenting behaviors*
- What is the benefit of the interventions relative to the cost?
- What is the mechanism through which the outcomes of interest are affected?
  - What is the impact of Save the Children's training on the service delivery and outreach of health workers?
  - Do service providers' deliver the program as intended?
  - What is the impact of the early childhood stimulation program on mothers' knowledge of early childhood practices?

Power calculation

	Unit Level 1
	Power (κ)
	0.8
Number of clusters	78
Cluster size	33
Intra-cluster correlation	.15
Significance level (α)	0.05
$t_{(1-\kappa)} + t_{\alpha}$	2.5
Sample size	2574
Minimum Detectable Effect	0.23

Evaluation design

- **Cluster Randomized Controlled Trail:** randomizes community clinics to treatment conditions and where the outcomes of interest are collected from households with small children.
- **Community Clinic:** 78 operational community clinics have been identified.
- **Intervention:**
  - All Community Clinic will implement NNS package
  - 50% of the CC (randomly selected) will implement the ECD intervention in addition to NNS package

Treatment	Control
National Nutrition Services (NNS) + ECD	National Nutrition Services (NNS)

Evaluation design cont..

- **Population and Study sample:**
  - **33 Households (<18 Months children) in the CC will be randomly selected for the study**
  - **24 months intervention**
  - **The same children will be evaluated at end line**
- **Study period: May-June, 2013 to May-June, 2015**

Condition	Treatment	Control
# of Community Clinic	**39 community clinics**	**39 community clinics**
# of Household in each CC	**33 Household (<18 months)**	**33 Household (<18 months)**
# of total Households	**1,287 households**	**1,287 households**

How do we see impact ?

- Comparing families in community clinic areas that **do and do not get PROGRAM**
- … but are **otherwise similar.**

- We use **randomization** to make certain both groups are similar
- Randomization is like a **lottery**

How do we randomize?

- **It stratifies by union to make sure that all unions will have both treated and control clinics.**
- **In unions with an even number of clinics, half are randomly assigned to treatment and half to the control condition.**
- **If a given union has an odd number of clinics, say N, half of (N-1) clinics are randomly assigned to treatment, the other half to control, and the treatment condition for the remaining clinic is randomly assigned between the union under consideration and another union with an odd number of clinics.**

Reference

Thank You