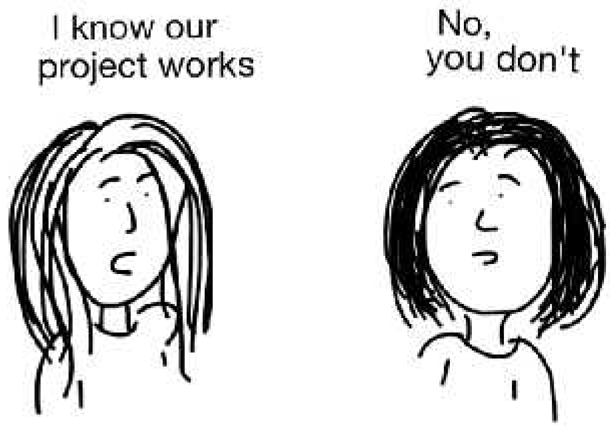
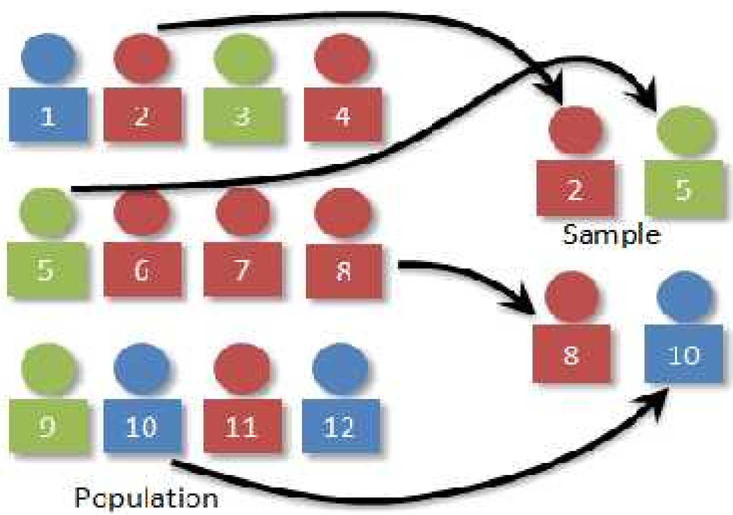
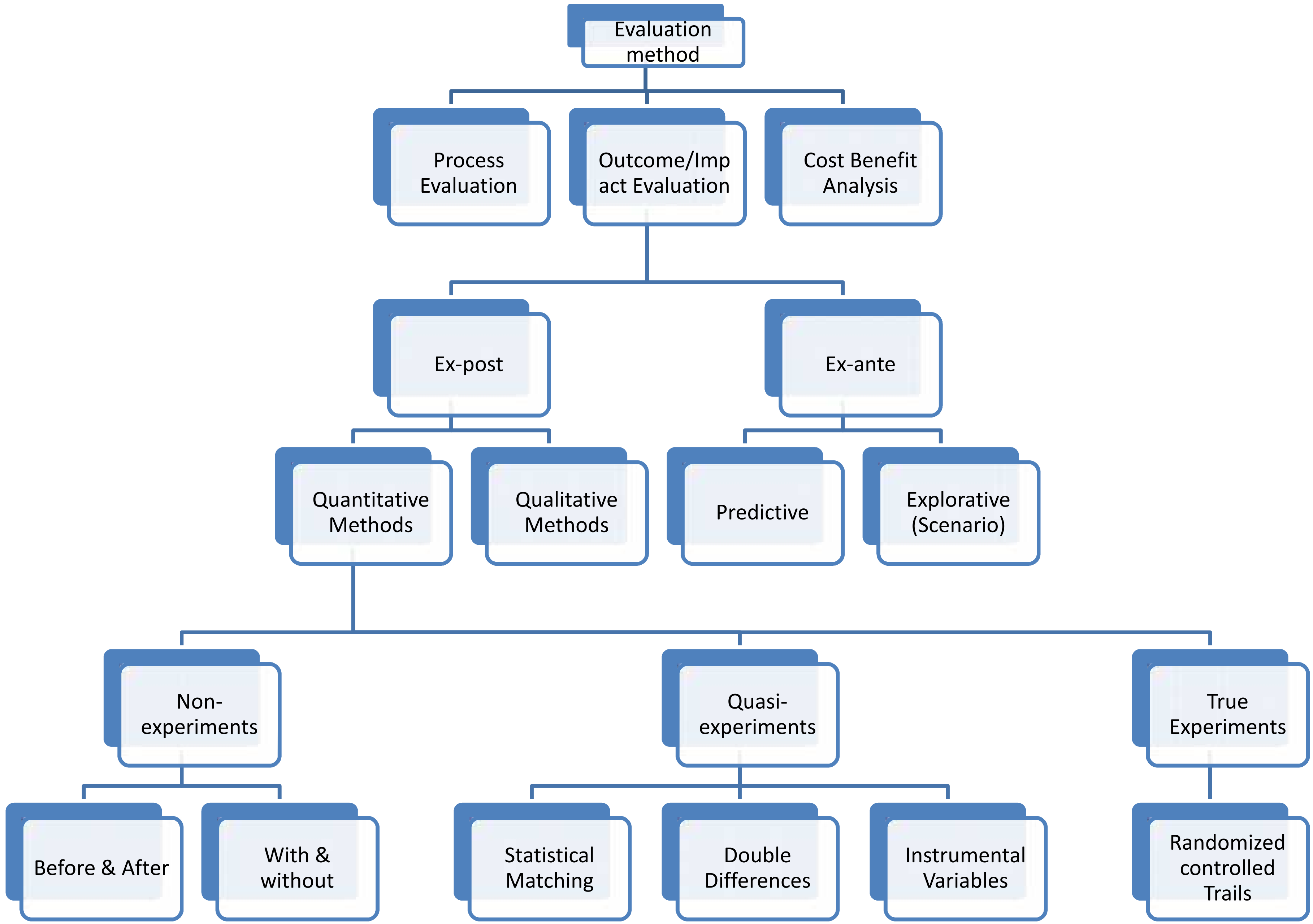
Randomisation and
Sampling Frames for Household Surveys: Guidelines



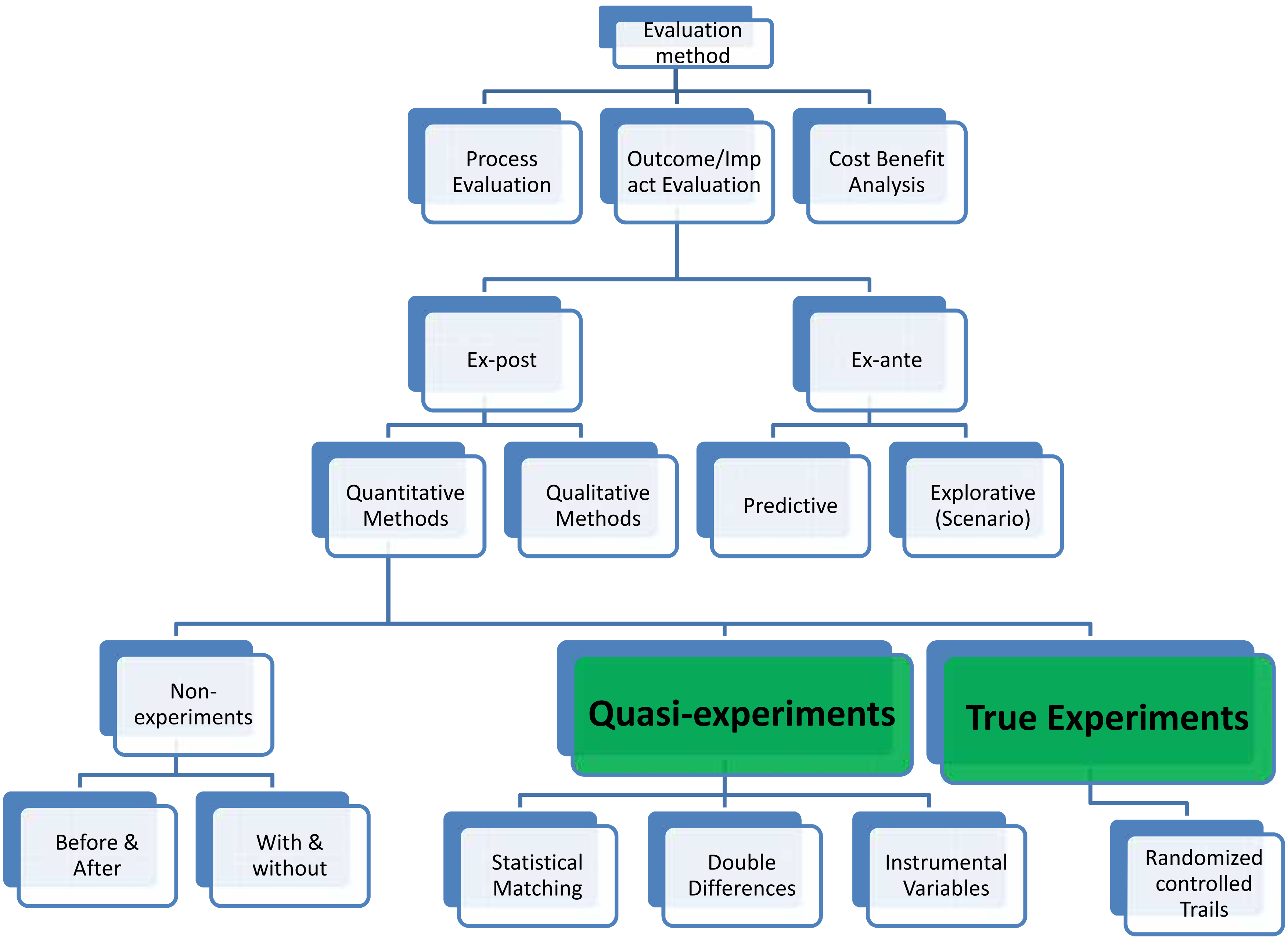


Srinivasulu Rajendran

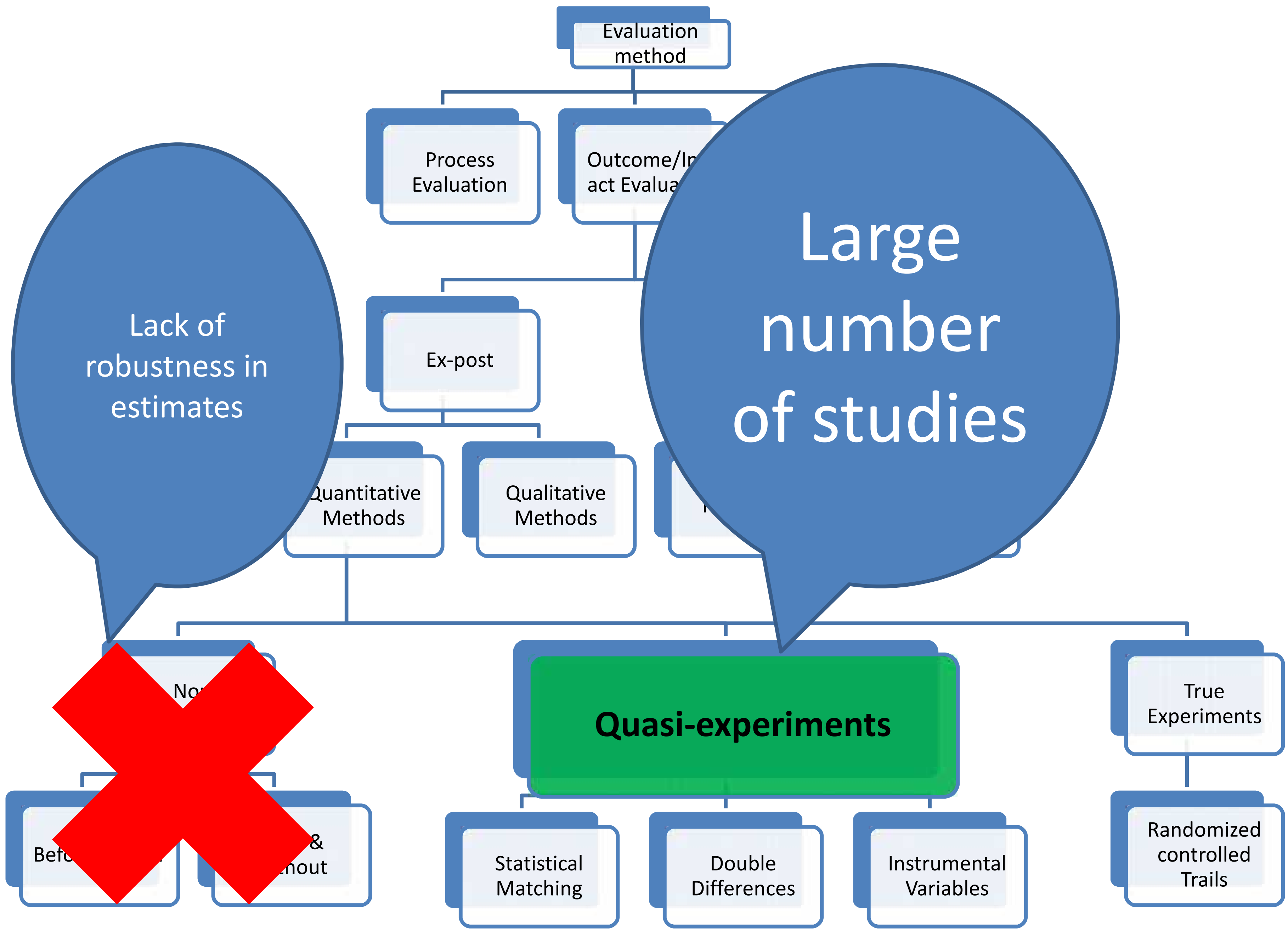
Monitoring, Learning, and Evaluation Community of Practice Meeting: 27 – 29 April, 2016 , Kigali, Rwanda



Source: Schreinemachers et al, 2014



Source: Schreinemachers et al, 2014



Source: Schreinemachers et al, 2014

Reasons for Chosen Quasi-experimental Method

- Size of the project and time duration
- Ethical issues associated with RCT
- Nature of project set-up

Structure of Presentation

- What and why Randomisation
- How to randomize?
- What is impact of the programme?
- Why do we need sampling?
- Type of sampling procedure
- Impact of the intervention
- Relevance of Sample size
- Testing Hypothesis – Types of errors
- Power calculation

What is Randomisation

- Rule: *all social programs should have fair and transparent rules for program assignment.*
- *How:* When a program is assigned at random over a large eligible population, we can generate a robust estimate of the counterfactual.
- allocating scarce resources among equally deserving populations turns out to be giving everyone who is eligible an equal opportunity to participate in the program.
- These randomized selection methods not only provide
- program administrators with a fair and transparent rule for allocating scarce resources *among equally deserving populations,* but also represent the strongest methods for evaluating the impact of a program.

Why Randomisation?

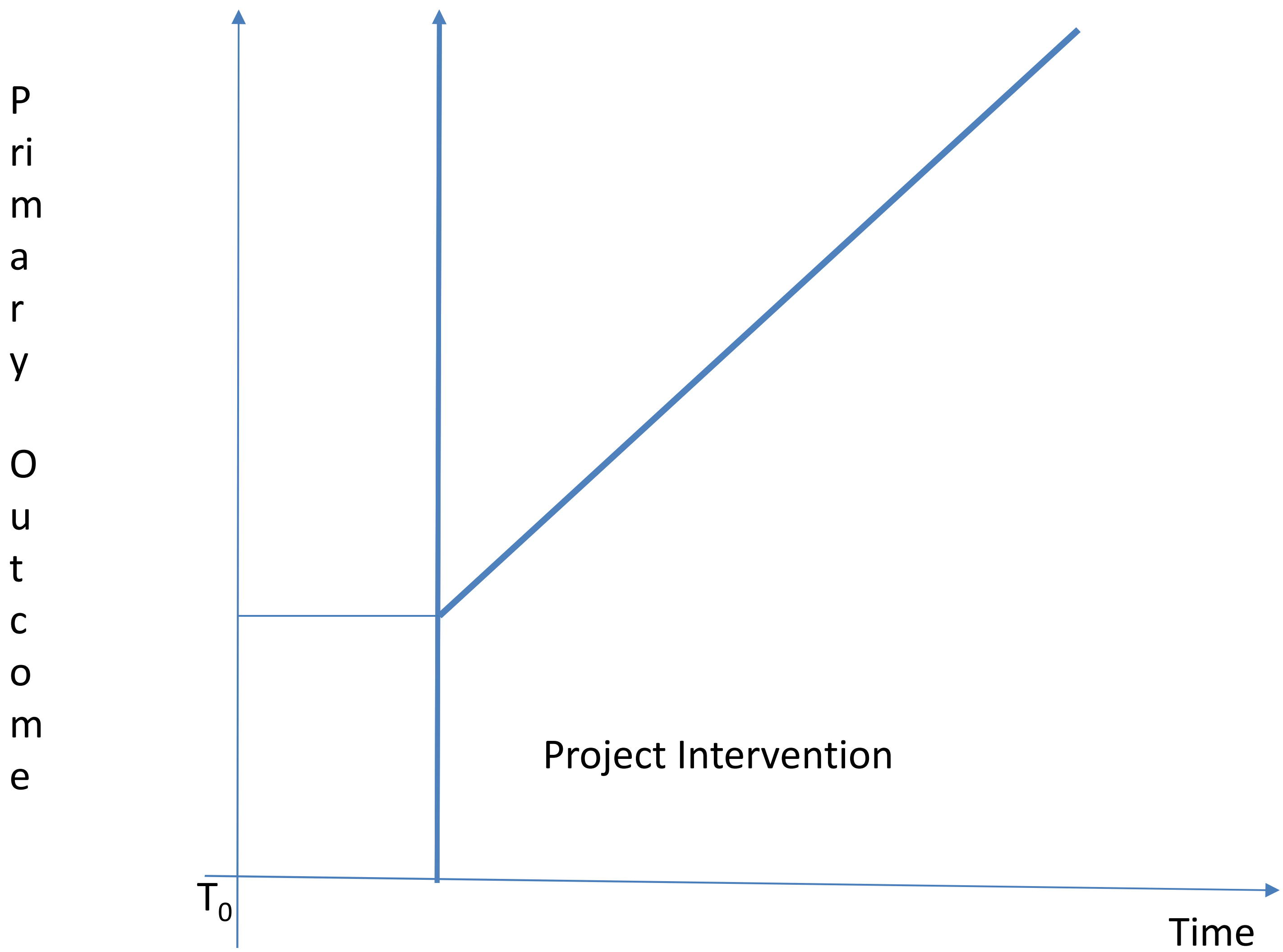
- Because members of the groups (treatment and control) do not differ systematically at the outset of the experiment,
- any difference that subsequently arises between them can be attributed to the program rather than to other factors.

How to randomize?

Methodologically, randomized trials are the best approach to estimate the effect of a program

1. Strongly disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly agree

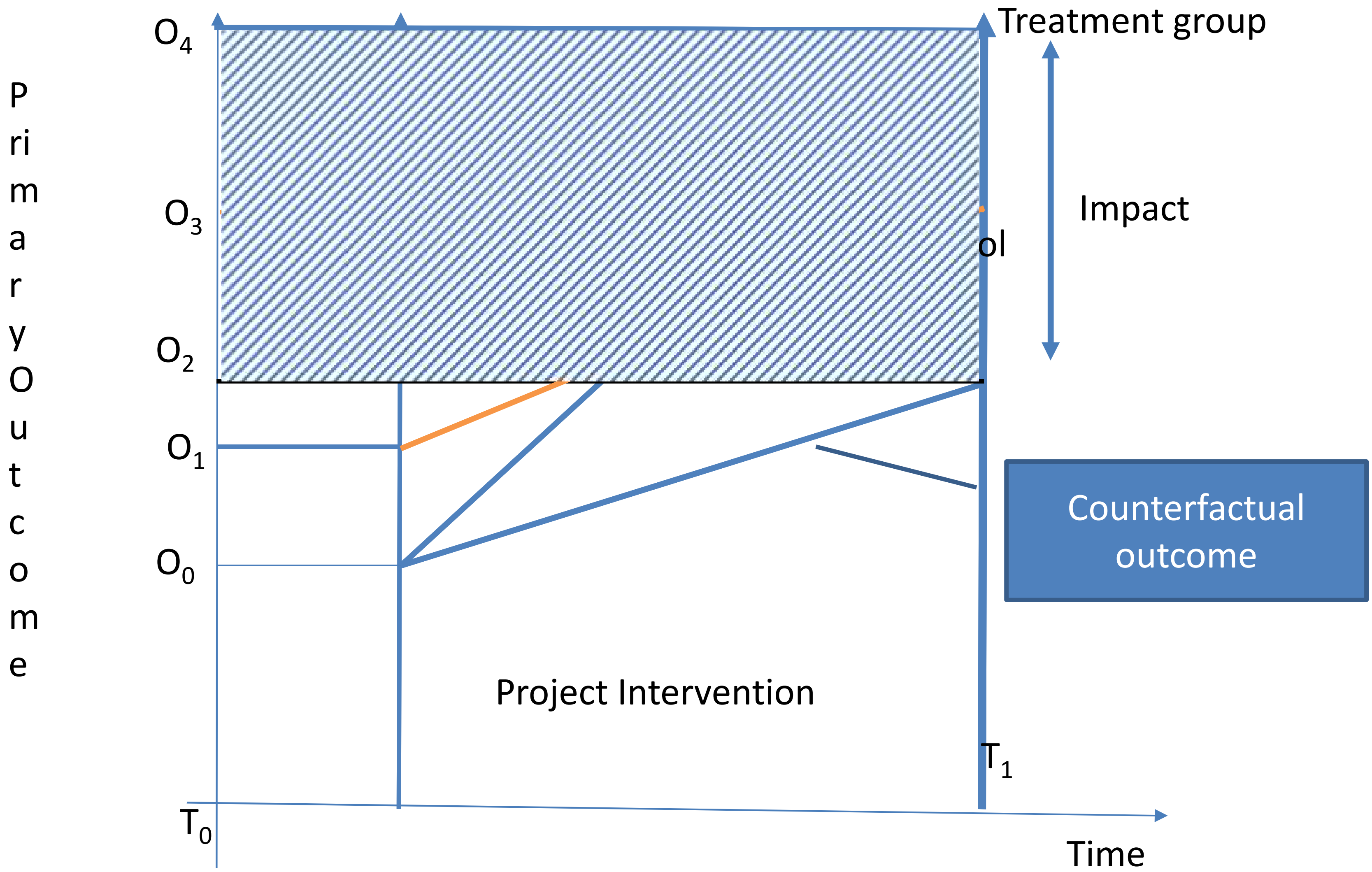
What is the impact of this program?



What is the impact of this program?

1. Positive
2. Negative
3. Zero
4. Not enough info

What is the impact of this program?



What is counterfactual?

- The counterfactual represents the state of the world that program participants would have experienced in the absence of the program (i.e. had they not participated in the program)

- Problem: Counterfactual cannot be observed

- Solution: We need to “mimic” or construct the counterfactual

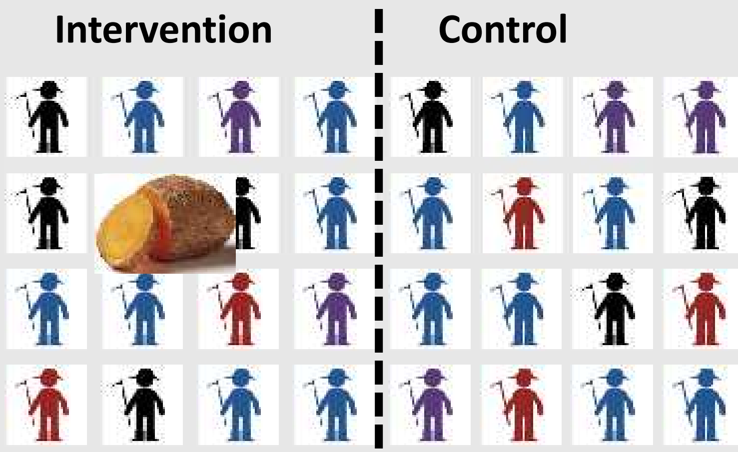
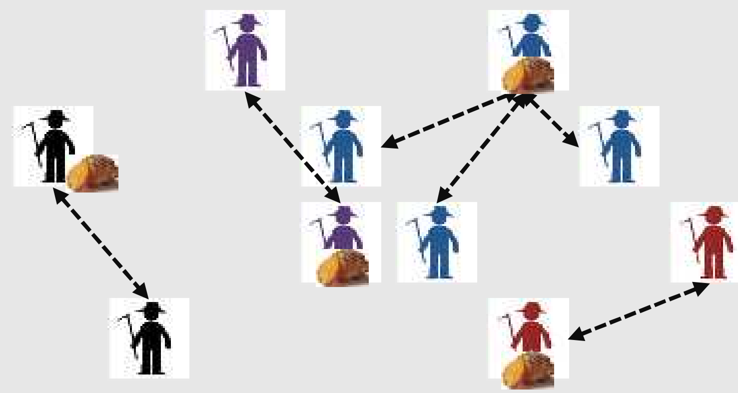
Constructing Counterfactual

- Usually done by selecting a group of individuals that did not participate in the program
- This group is usually referred to as the control group or comparison group
- How this group is selected is a **key decision** in the design of any impact evaluation

Impact of the programme intervention

- CIP would like to see the impact of programme intervention introduced by CIP – treatment and control group
- Carry out impact assessments
- Outcomes – yield and dietary diversity

Impact evaluation (IE)

Experimental	**Observational**
We allocate eligible subjects randomly to intervention and control	We compare adopters with non-adopters
	
• Collect data on impact indicators and household characteristics before and after the intervention	• Collect data on impact indicators and household characteristics before and after the intervention • Use statistical methods to construct an valid counterfactual

Challenges in identifying impacts

- Note that we compare these two groups at endline. We can choose to include the change in their outcomes from the baseline.

- BUT how do we know the impact is not due to random chance from sampling?

- Creating a statistically equal groups before the study.

Q&A

- Did the our programme intervention improve yield level and dietary diversity outcomes?
- Did the programme have a statistically significant impact on outcomes?

Relevance of Sample Size (Magruder, 2016)

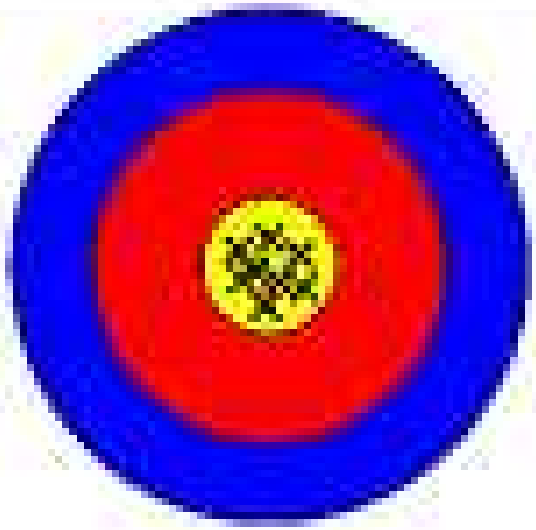
- Randomization removes the bias (ensures accuracy) but it does not remove noise
- We can limit random chance through precision
- Precision is controlled through sample size
- By increasing our sample size we increase the precision of our distribution and decrease the likelihood of random chance.

Types of Sampling

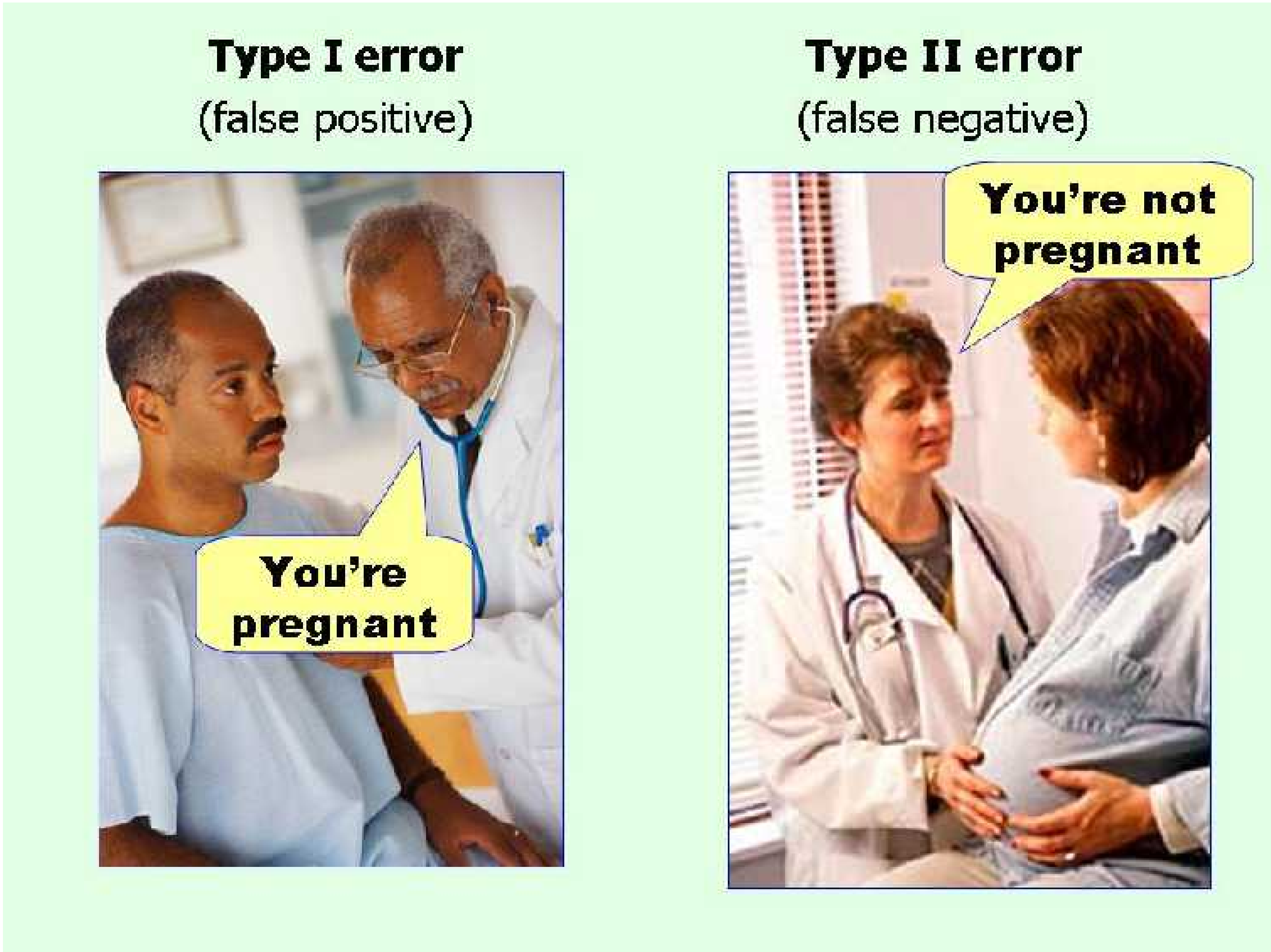
- Probability Sampling – Simple Random Sampling, Stratified Random Sampling, Multi-Stage Sampling
- Non-probability Sampling – Accidental and purposive sampling (Modal Instance sampling, Expert Sampling, Quota Sampling, Heterogeneity sampling and Snowball Sampling)

Q&A

- The difference between nonprobability and probability sampling is that nonprobability sampling does not involve *random* selection and probability sampling does. Does that mean that nonprobability samples aren't representative of the population?
- A) yes b) No, not necessary c) no other options

	Accurate	Inaccurate (systematic error)
Precise		
Imprecise (reproducibility error)		

Types of Errors



What is Power of hypothesis test?

- **Type II Error:** Failing to reject the null hypothesis (concluding there is no difference), when indeed the null hypothesis is false.
- **Power of hypothesis test:** The probability that we will **detect an effect** (reject the null hypothesis) if our intervention has a measureable effect (when alternative hypothesis is true). This what we want.
- Power = 1 – Probability of type of II error

Q&A

- What are the factors that change the proportion of the research hypothesis that is shaded i.e the proportion that falls to the right (or left) of the null hypothesis curve?
- Understanding this helps us design more powerful experiments.

Power: main ingredients

- **Sample Size**
- Effect Size
- Variance
- Proportion of sample treatment and control
- Clustering

Larger sample size = more power to detect

- Ex: we want to know average yield of X crops in the study region,
- If we randomly pick one sample in the study region, we might pick a farmer who got high yield.
- If we randomly pick up more farmers, even if we pick the farmer who produce high yield, he will be balanced out by the other random selections.

Measures of Noise and Effect of sample size

- **Standard Deviation**: A measure of dispersion from the mean of the underlying population from which subjects are sampled.
- **Standard error**: A measure of precision of an estimated size of an effect (the larger our standard error, the less precise our estimate)
- When there is a rise in sample size, SE goes down, which will improve our estimates.

Power: main ingredients

- Sample Size
- **Effect Size**
- Variance
- Proportion of sample treatment and control
- Clustering

Larger effect = more power to detect

- A device detects all farmers who uses high yield variety (X variety).
- Power to detect farmer who use P variety: Under 10%
- Power to detect farmer who use Q variety: Under 1%
- Power to detect farmer who use R variety: 0%
- Power to detect farmers who use X variety: 100%
- The farmer who use X variety (effect size) we care about, the more power we have (and the less we need)

Q&A

- Do you think effect size equals distance between average scores in treatment and control group?
- A) Yes
- B) No
- C) No other option

- What effect size should you use when designing your experiments?
- A) smallest effect size that is still cost effective
- B) largest effect size you expect your programme to produce
- C) both
- D) Neither

Picking an Effect Size to choose sample

We can guess an effect size using

- Economics
- Past data on the outcome of interest or even past evaluations

What is the smallest effect that should justify the program to be adopted?

- Cost of this program vs. the benefits it brings
- Cost of this program vs. the alternative use of the money

Power: main ingredients

- Sample Size
- Effect Size
- **Variance**
- Proportion of sample treatment and control
- Clustering

More variance = Less power to detect

- Imagine the following intervention: Giving away ten bags of vines

  - In this example, this program has a large effect on ALL poor people, and no effect on ALL rich people.

- Low Variance: If our population is all poor, we only need to sample one person to see the true effect of giving away vines
- High Variance: If our population is half poor, and half rich (high variance) and we randomly sample twenty people, what happens if only 5 are poor?

Q&A

- What are typical ways to reduce the underlying (population) variance?
- A) include covariates
- B) increase the sample size
- C) do a baseline survey
- D) all of the above
- E) A and B
- F) A and C

Variance

- There is sometimes very little we can do to reduce the noise
- The underlying variance is what it is
- We can try to “absorb” variance:

- using a baseline
- controlling for other variables
- In practice, controlling for other variables (besides the baseline outcome) buys you very little

What should I do if I don’t have enough power?

- A) continue the study anyway
- B) continue the study and note that you are underpowered in the results
- C) stop the study
- D) increase your expected effect size

Big Mistake! Underpowered

- Common danger: picking effect sizes that are too optimistic— the sample size may be set too low to detect an actual effect
- Example: Evaluators believe a program will increase yield level by 15 percentage points.
- They survey enough farmers to see increases of 12 percentage points or more.
- The program increased yield level by 10 percentage points, but they missed that entirely due to lack of power
- They report the program had no statistically significant effect, even though it actually had an impact!

Power: main ingredients

- Sample Size
- Effect Size
- Variance
- **Proportion of sample treatment and control**
- Clustering

- If samples are 50% and 50% : Equal split gives distributions that are the same “fatness”
- If samples are 75% and 25%: Uneven distributions, not efficient, i.e., less power

Power: main ingredients

- Sample Size
- Effect Size
- Variance
- Proportion of sample treatment and control
- **Clustering**

Clustered design: definition

- In sampling (Quasi-experiments):

- When clusters of individuals (e.g. Communities, farmers associations, etc) are randomly selected from the population, before selecting individuals for observation

- In randomized evaluation (i.e., RCT):

- When clusters of individuals are randomly assigned to different treatment groups

Reason for adopting cluster randomization

- Need to minimize or remove contamination
  Example: In the deworming program, schools was chosen as the unit because worms are contagious
- Basic feasibility considerations

Example: Some programs would not have been politically feasible if some interventions were introduced.

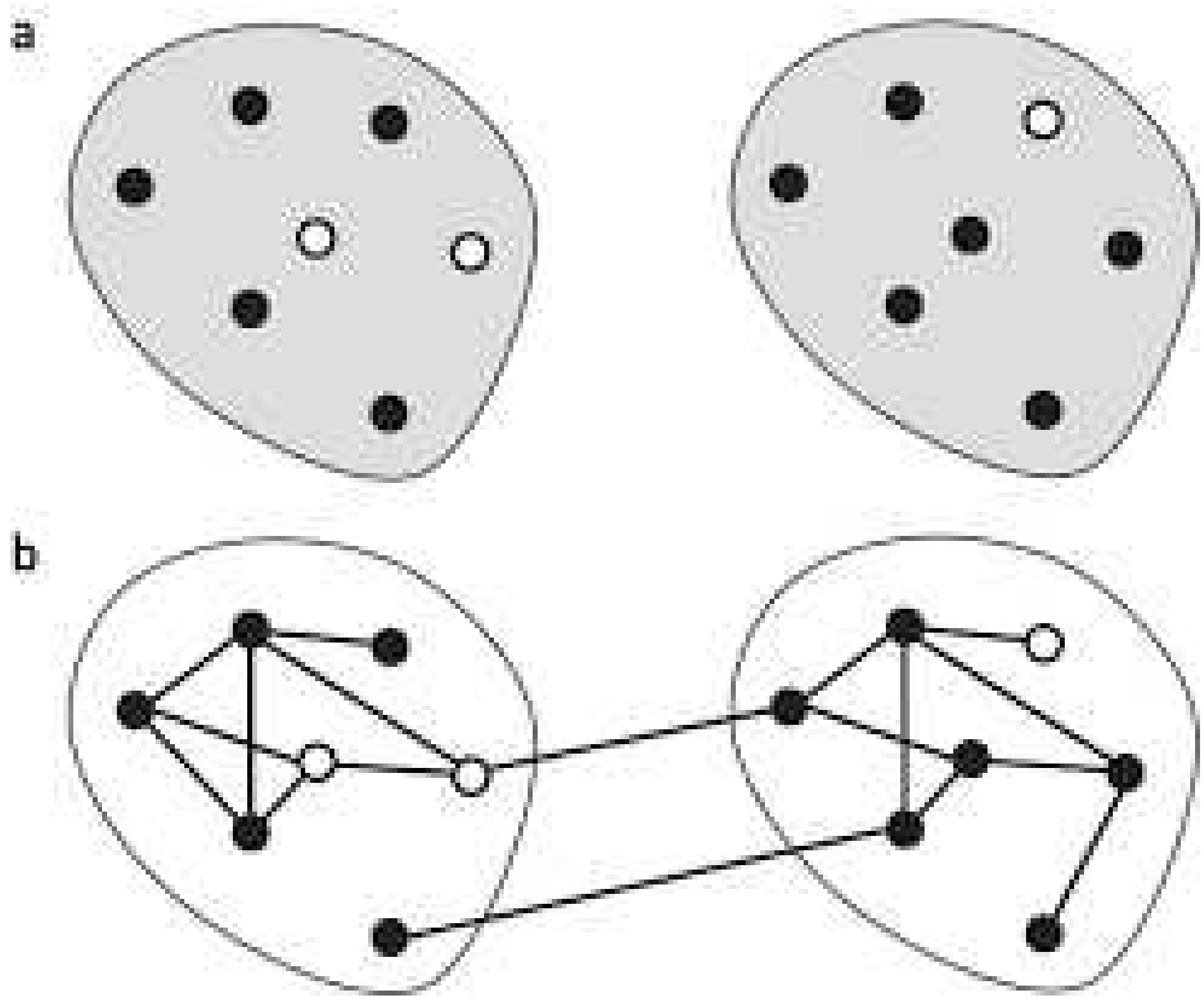
- Only natural choice

Example: Any intervention that affect an entire village (e.g. training etc).

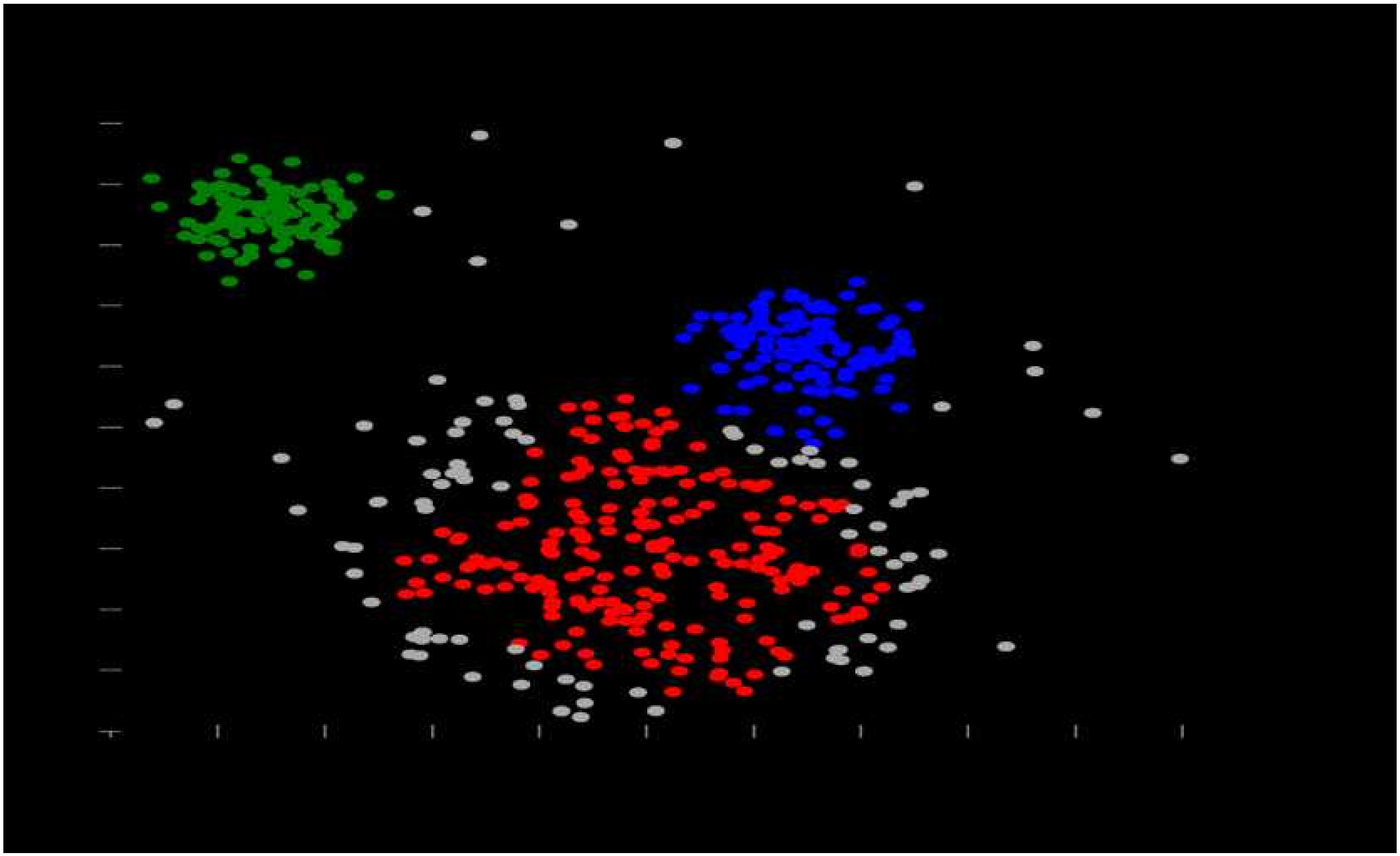
Clustered design: intuition

- You want to know how close the upcoming national elections will be
- Method 1: Randomly select 50 people from entire Indian population
- Method 2: Randomly select 5 families, and ask ten members of each family their opinion

Low intra-cluster correlation (ICC)



HIGH intra-cluster correlation



Q&A

- All large farmers lives in one village, middle farmers lives in another village and smallholders lives in some other village. ICC (ρ) will be ……
- A) high
- B) low
- C) no effect on rho
- D) don't know

Clustered Design: Intuition

- The outcomes within a family are likely correlated. Similarly with children within a school, families within a village etc.
- Each additional individual does not bring entirely new information
- At the limit, imagine all outcomes within a cluster are exactly the same: effective sample size is number of clusters, not number of individuals
- Precision will depend on the number of clusters, sample size within clusters and the within cluster correlation.

Q&A

- If ICC is high, what is a more efficient way of increasing power?
- A) include more clusters in the sample
- B) include more people in cluster
- C) both
- D) don’t know

Wrap up on Power

- Always must conduct power calculations before conducting a randomized evaluation to ensure sample size is big enough to detect anticipated effects.
- If you do not have a sample big enough, either expand your sample or do not do the study!
- You learn nothing from an underpowered randomized evaluation.

Asante Saana!!!