

ABDUL LATIF JAMEEL
Poverty Action Lab



clear



TRANSLATING RESEARCH INTO ACTION

Impact Evaluation Methods

April 9, 2012

Diva Dhar

CLEAR Regional Center

J-PAL South Asia

povertyactionlab.org

Impact evaluation methods

- a. Pre-Post
- b. Simple Difference
- c. Differences-in-Differences
- d. Multivariate Regression
- e. Statistical Matching
- f. Instrumental Variables
- g. Regression Discontinuity
- h. Randomized Evaluations

Example: Pratham's Balsakhi Program



What was the problem?

- Many children in 3rd and 4th standard were not even at the 1st standard level of competency
- Class sizes were large
- Social distance between teacher and some of the students was large

Context and Partner

- 124 Municipal Schools in Vadodara (Western India)
- 2002 & 2003: Two academic years
- ~ 17,000 children
- Pratham – “Every child in school and learning well”
- Works with most states in India reaching millions of children

Proposed solution

- Hire local women (balsakhis) from the community
- Train them to teach remedial competencies
 - Basic literacy, numeracy
- Identify lowest performing 3rd and 4th standard students
 - Take these students out of class (2 hours/day)
 - Balsakhi teaches them basic competencies

Possible outcomes

Pros

- Reduced social distance
- Reduced class size
- Teaching at appropriate level

- Improved learning for lower-performing students
- Improved learning for higher-performers

Cons

- Less qualified
- Teacher resentment
- Reduced interaction with higher-performing peers

- Increased gap in learning
- Reduced test scores for all kids

What was the impact?

We conduct a test at the end

- Balsakhi students score an average of 51%



What can we conclude?

1 - Pre-post (Before vs. After)

Average change in the outcome of interest
before and after the programme

1 - Pre-post (Before vs. After)

- Look at average change in test scores over the school year for the balsakhi children



1 - Pre-post (Before vs. After)

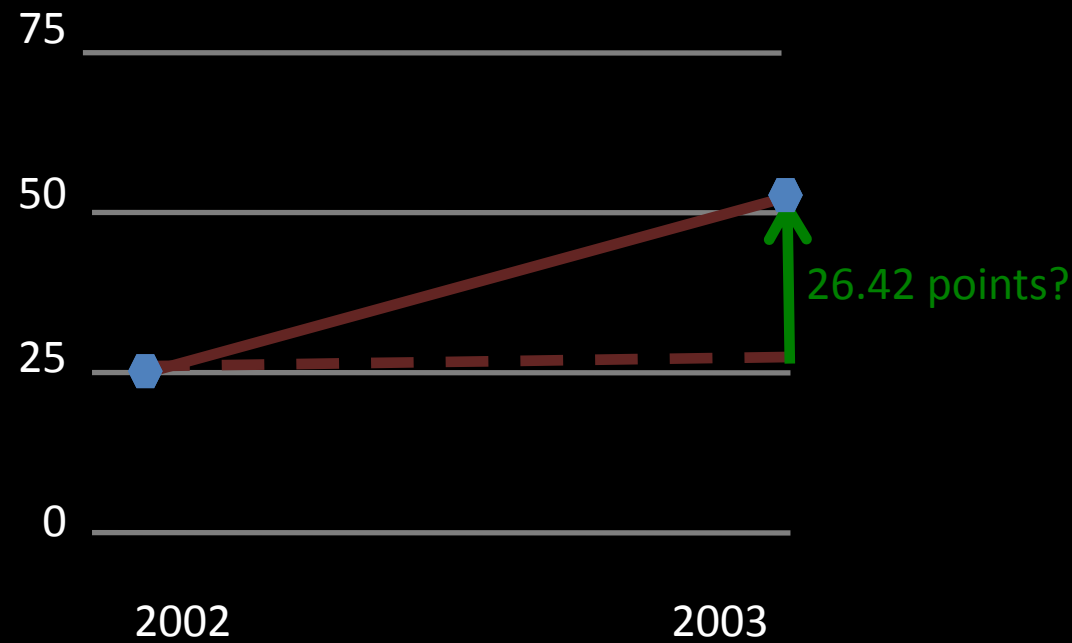
Average <u>post-test</u> score for children with a balsakhi	51.22
Average <u>pretest</u> score for children with a balsakhi	24.80
Difference	26.42

- **QUESTION:** Under what conditions can this difference (26.42) be interpreted as the impact of the balsakhi program?

Pre-post

Method 1: Before vs. After

Impact = 26.42 points?



What would have happened without the balsakhi program?

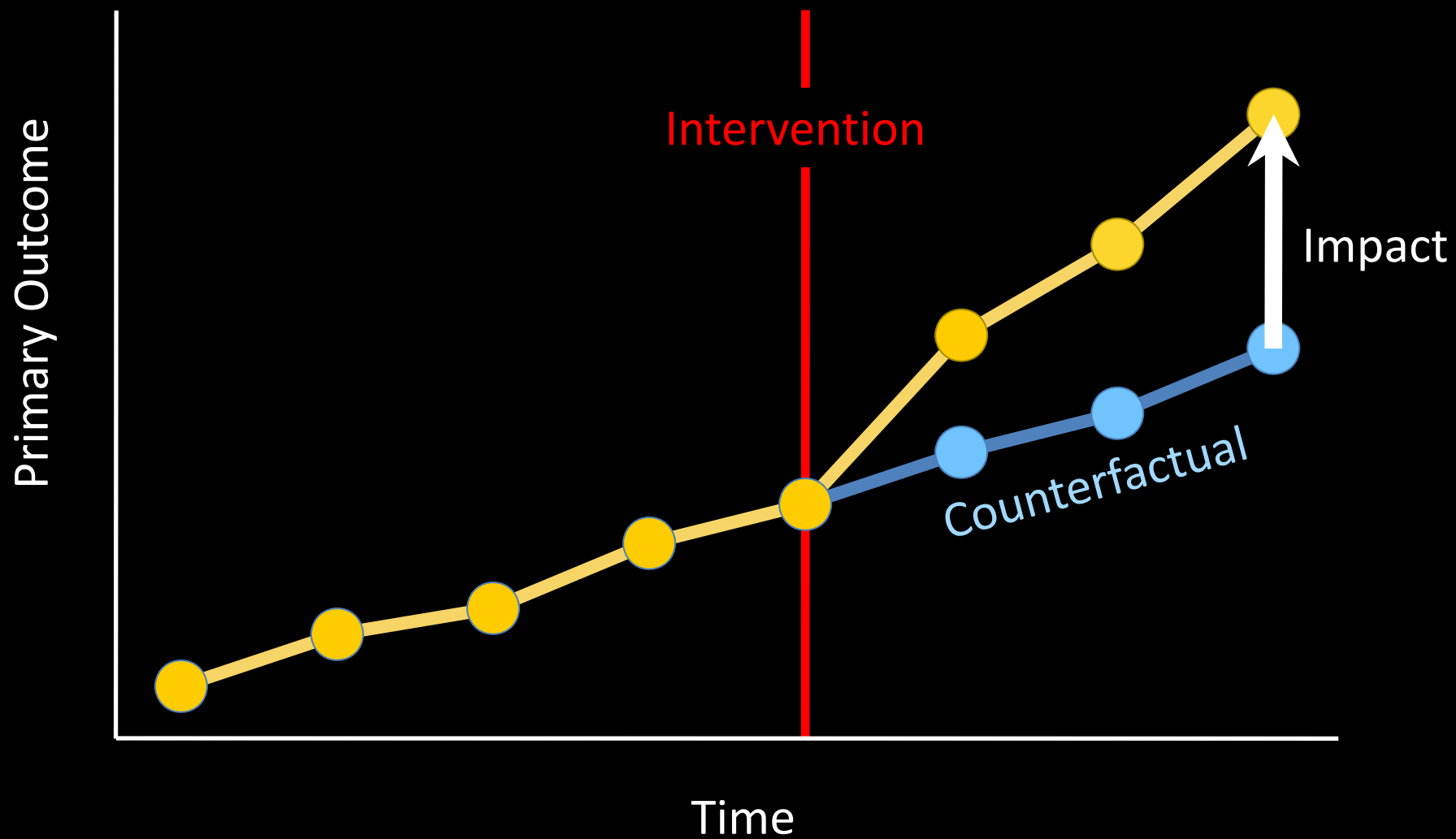
How to measure impact?

Impact is defined as a comparison between:

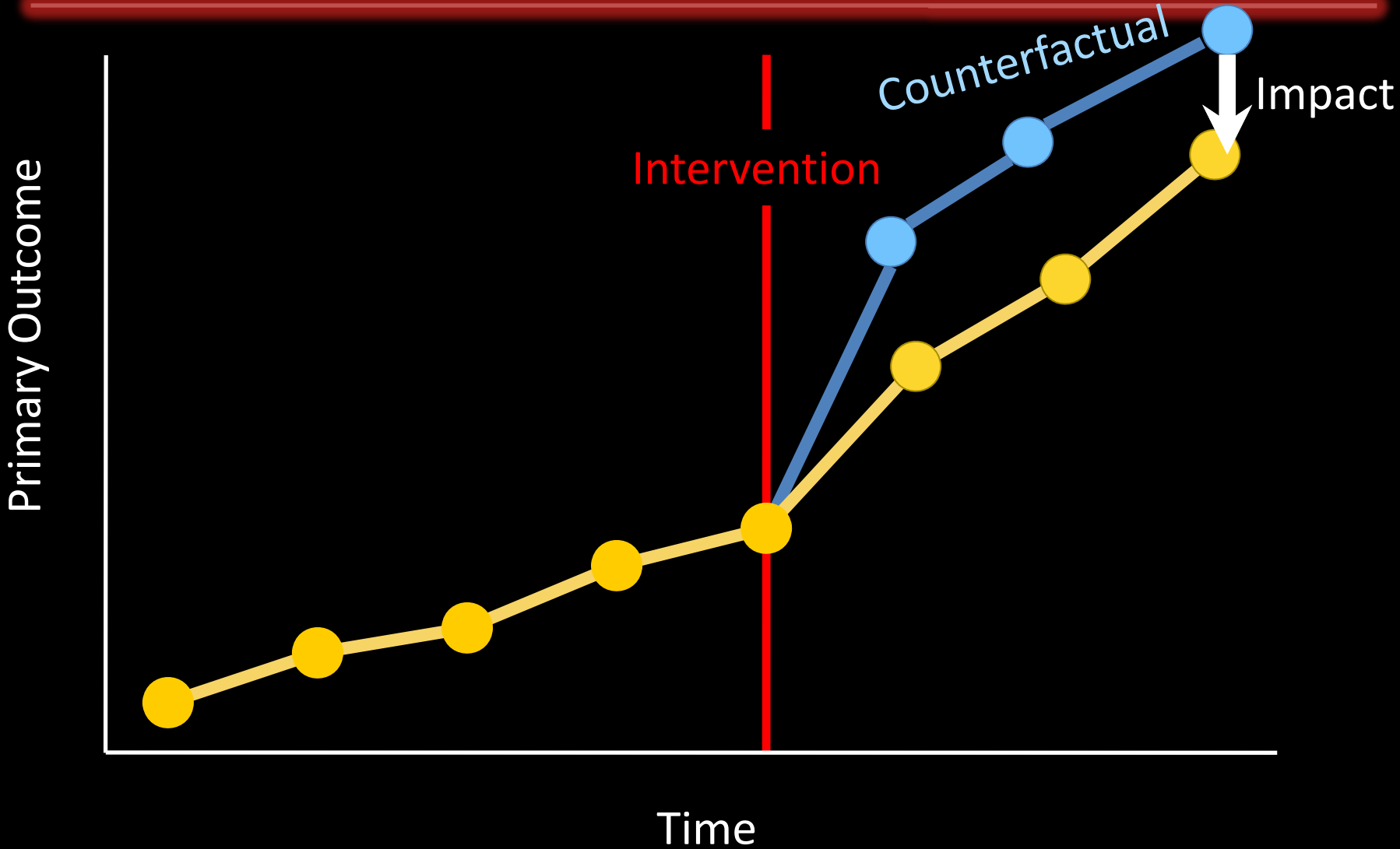
1. the outcome some time after the program has been introduced
2. the outcome at that same point in time had the program not been introduced

the "*counterfactual*"

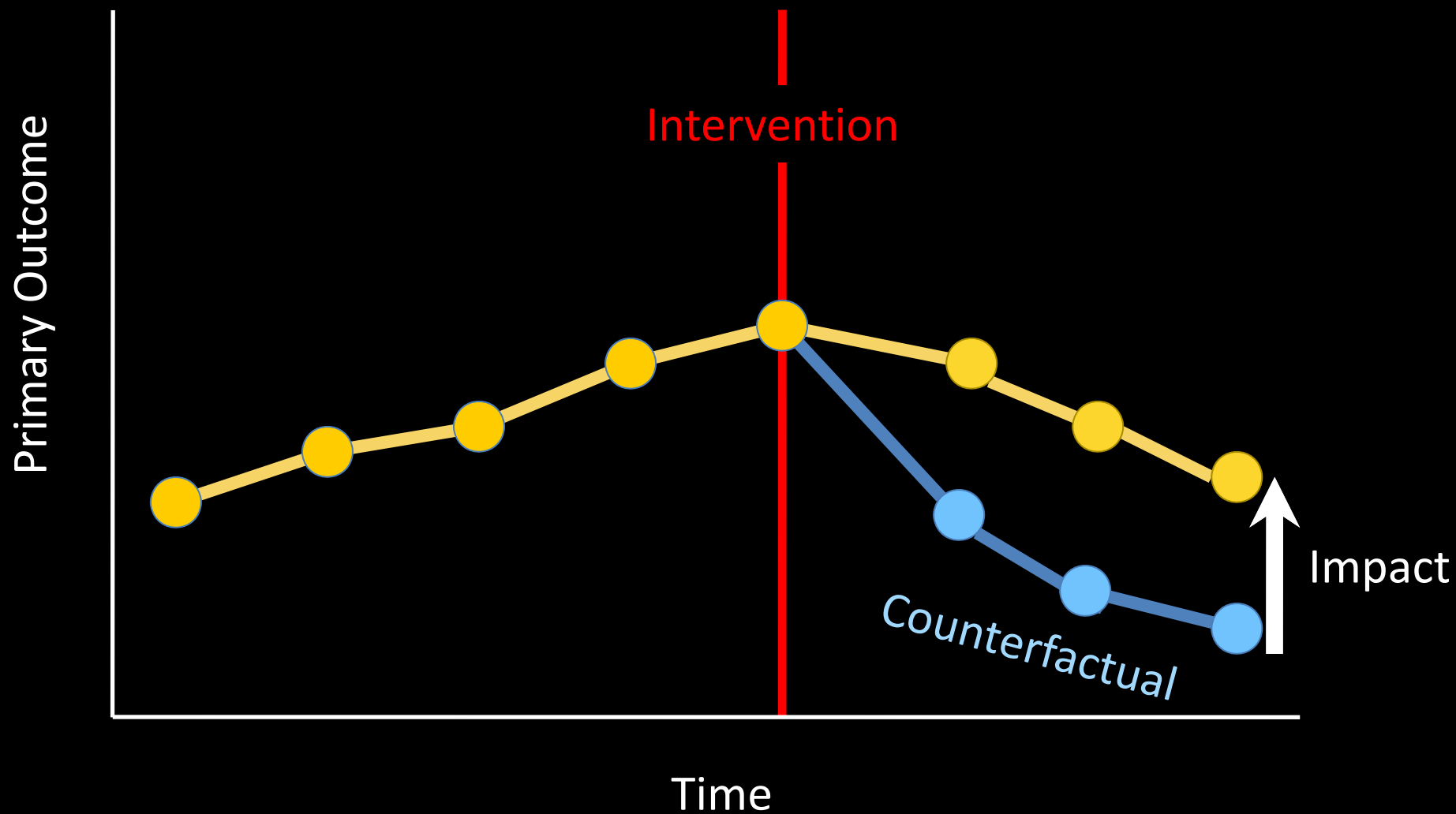
Impact: What is it?



Impact: What is it?



Impact: What is it?



Pre-Post:

- Limitations of the method: No comparison group, doesn't take time trend into account



What else can we do to estimate impact?

2 - Simple difference

A post- programme comparison of outcomes between the group that received the programme and a “comparison” group that did not

- **Example:**
 - programme is rolled out in phases leaving a cohort for comparison, even though the assignment of the program is not random

2 - Simple difference

Compare **post-program test scores** of...



Children who **got**
balsakhi

With
test
scores
of...



Children who **did not** get
balsakhi

2 - Simple difference

Average score for children with a balsakhi	51.22
Average score for children without a balsakhi	56.27
Difference	-5.05

- **QUESTION:** Under what assumptions can this difference (-5.05) be interpreted as the impact of the balsakhi program?

3 – Difference-in-Differences

Compare gains in test scores of...



Children who **got**
balsakhi

With
gains
in test
scores
of...



Children who **did not** get
balsakhi

3 – Difference-in-Differences (or Double Difference)

Comparison of outcome between a treatment and comparison group (1st difference) and before and after the programme (2nd difference)

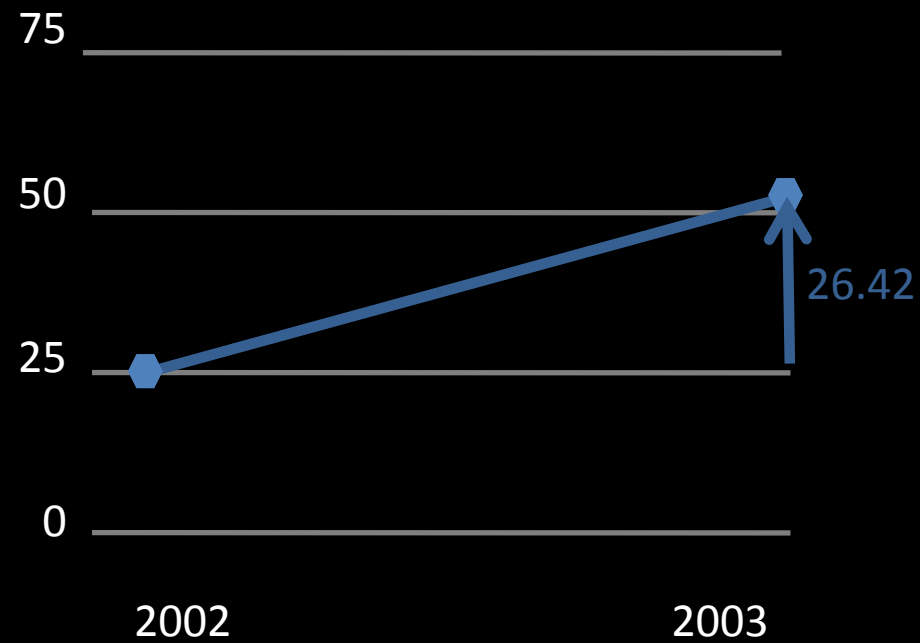
- Suitability:
 - programme is rolled out in phases leaving a cohort for comparison, even though assignment of treatment is not random

3 - Difference-in-differences

	Pretest	Post-test	Difference
Average score for children with a balsakhi	24.80	51.22	26.42

What would have happened without balsakhi?

Method 3: Difference-in-differences

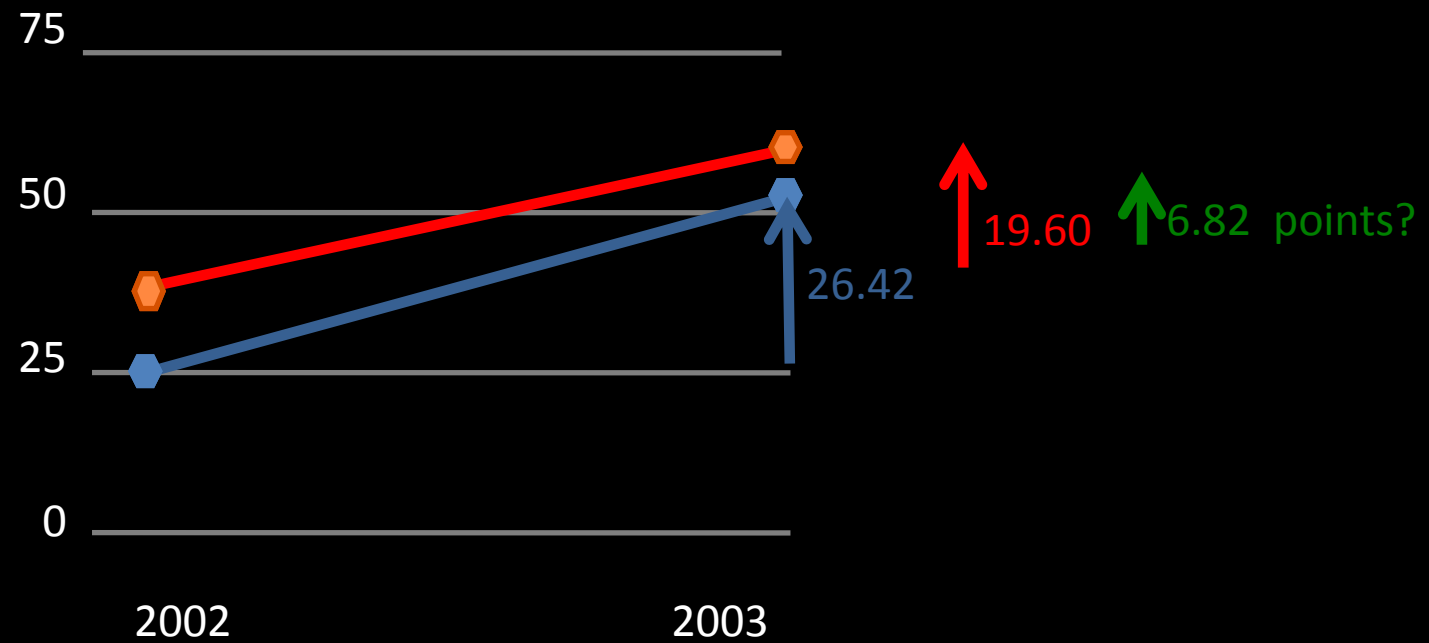


3 - Difference-in-differences

	Pretest	Post-test	Difference
Average score for children with a balsakhi	24.80	51.22	26.42
Average score for children without a balsakhi	36.67	56.27	19.60

What would have happened without balsakhi?

Method 3: Difference-in-differences

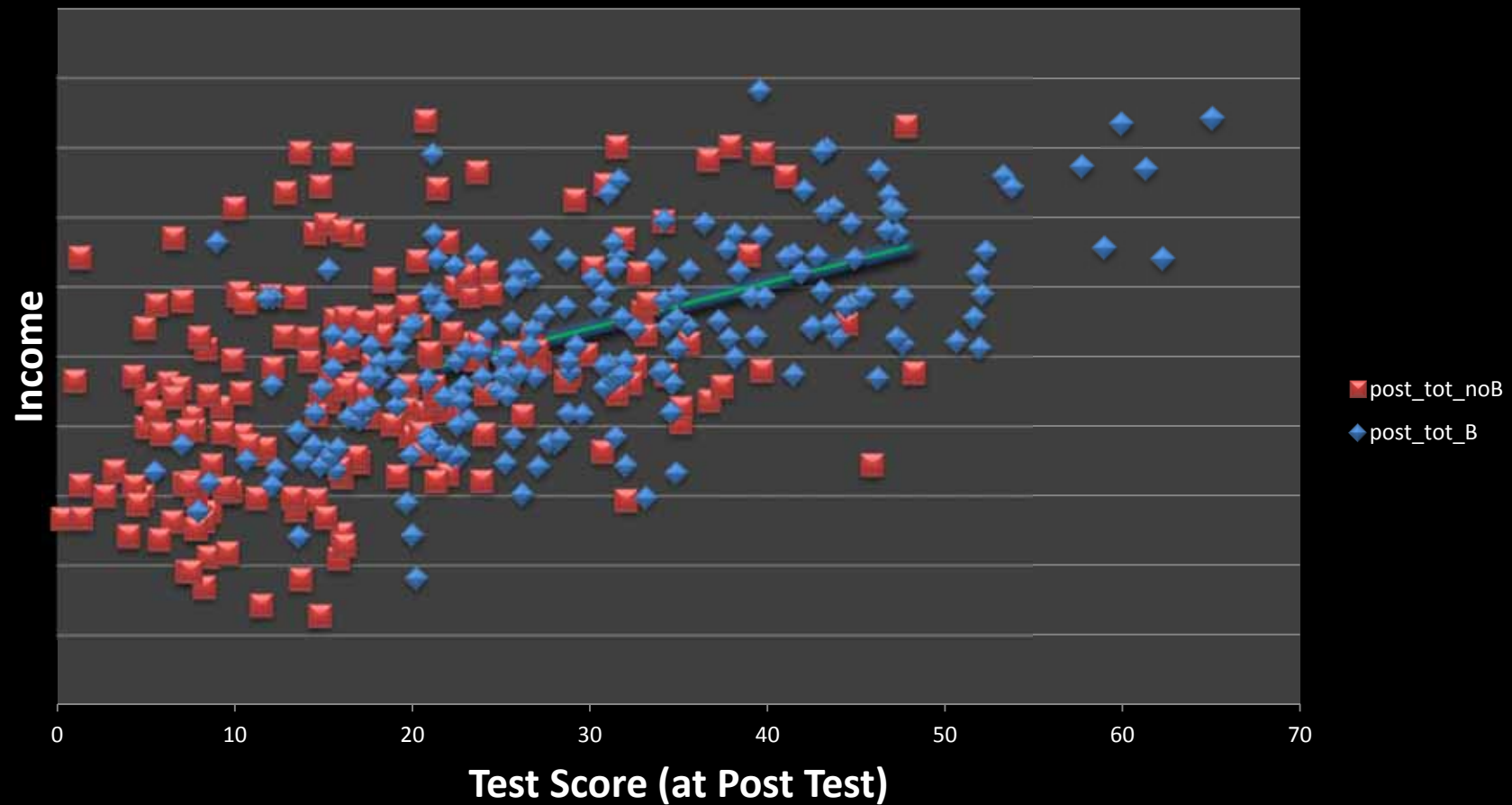


3 - Difference-in-differences

	Pretest	Post-test	Difference
Average score for children with a balsakhi	24.80	51.22	26.42
Average score for children without a balsakhi	36.67	56.27	19.60
Difference			6.82

- **QUESTION:** Under what conditions can 6.82 be interpreted as the impact of the balsakhi program?
- **Issues:**
 - failure of “parallel trend assumption”, i.e. impact of time on both groups is not similar

4 - Accounting for other factors



Impact of Balsakhi - Summary

Method	Impact Estimate
(1) Pre-post	26.42*
(2) Simple Difference	-5.05*
(3) Difference-in-Difference	6.82*
(4) Regression with controls	1.92

*: Statistically significant at the 5% level

5 – Other Methods

- There are more sophisticated non-experimental and quasi-experimental methods to estimate program impacts:
 - Multivariable Regression
 - Matching
 - Instrumental Variables
 - Regression Discontinuity
- These methods rely on being able to “mimic” the counterfactual **under certain assumptions**
- **Problem:** Assumptions are not testable

Constructing the counterfactual

- Counterfactual is often constructed by selecting a group not affected by the program
- Non-randomized:
 - Argue that a certain excluded group mimics the counterfactual.
- Randomized:
 - Use random assignment of the program to create a control group which mimics the counterfactual.

Randomized Evaluations

- Individuals, clients, firms, villages are *randomly* selected to receive the treatment, while other units serve as a comparison

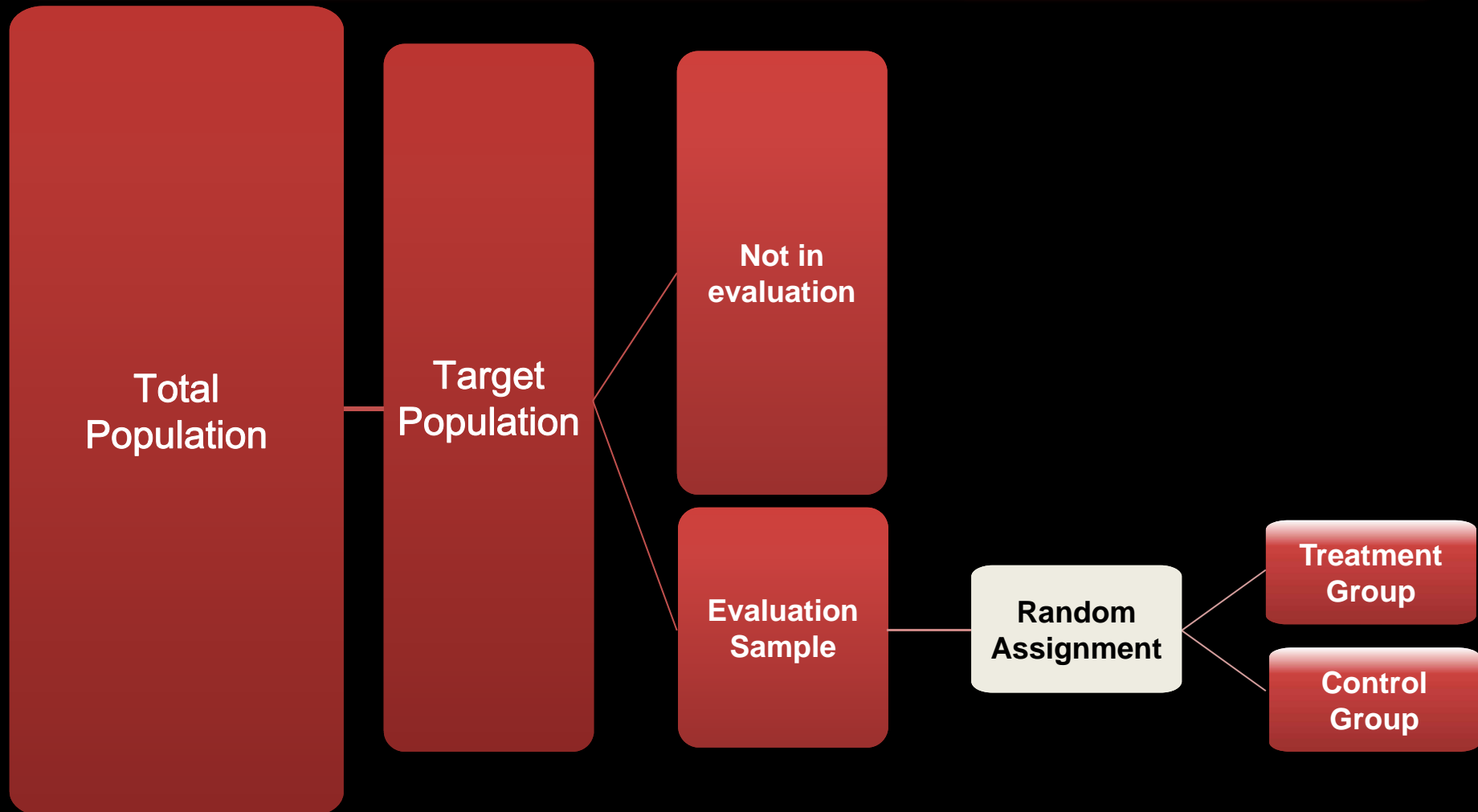
GROUPS ARE STATISTICALLY IDENTICAL BEFORE PROGRAM



ANY DIFFERENCES AT ENDLINE CAN BE ATTRIBUTED TO PROGRAM

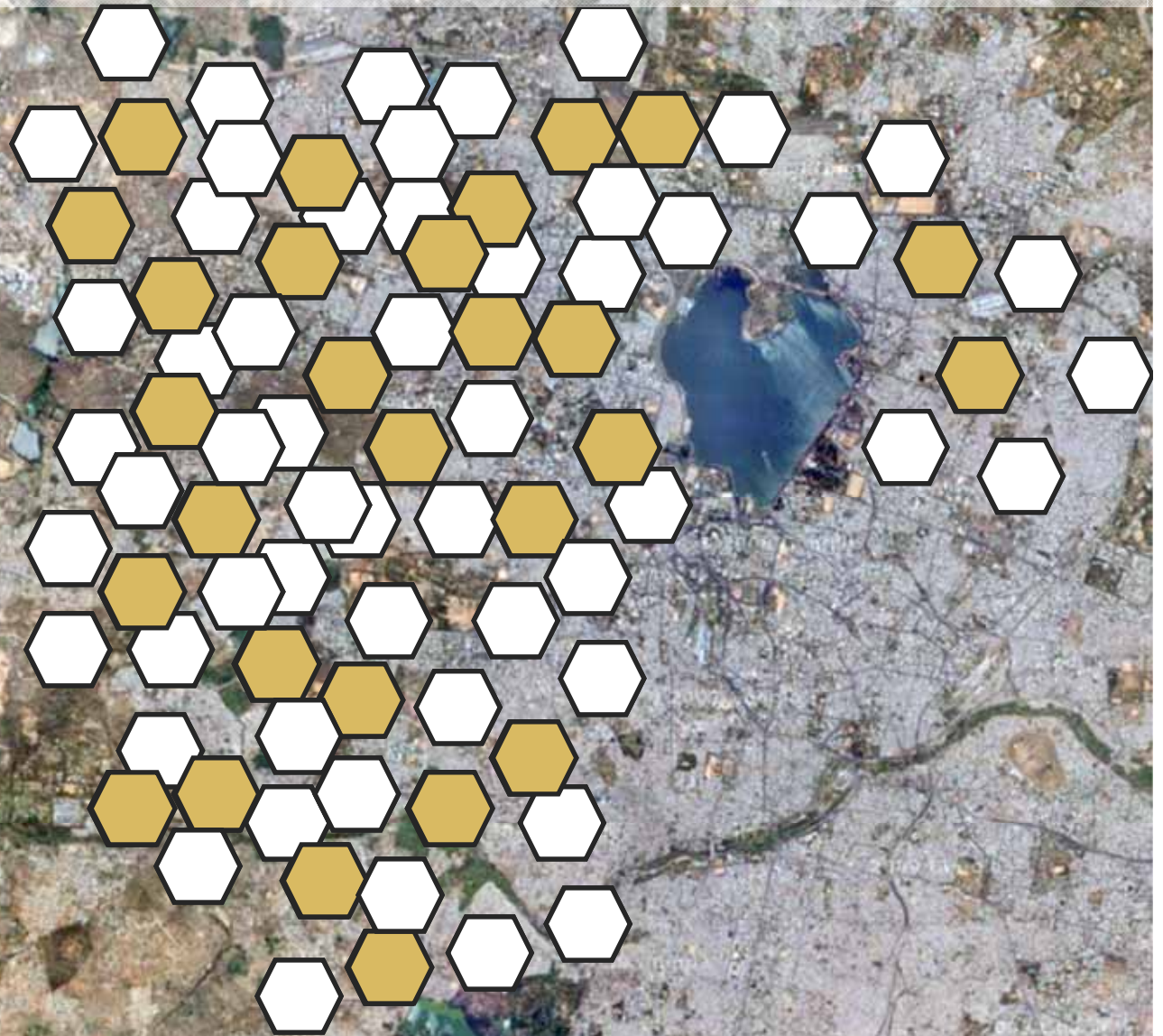


Basic set-up of a randomized evaluation



Random sampling and random assignment

Randomly
sample
from area of
interest

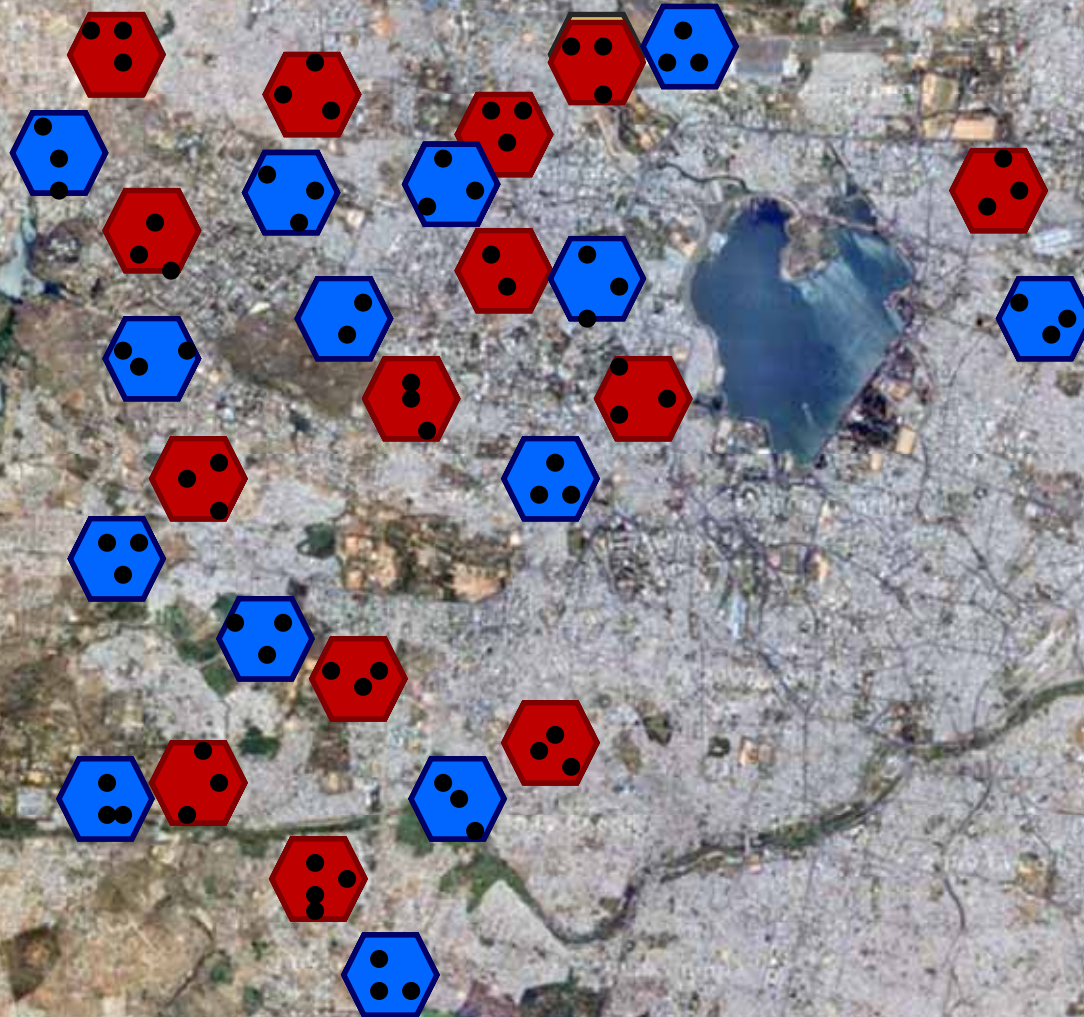


Random sampling and random assignment

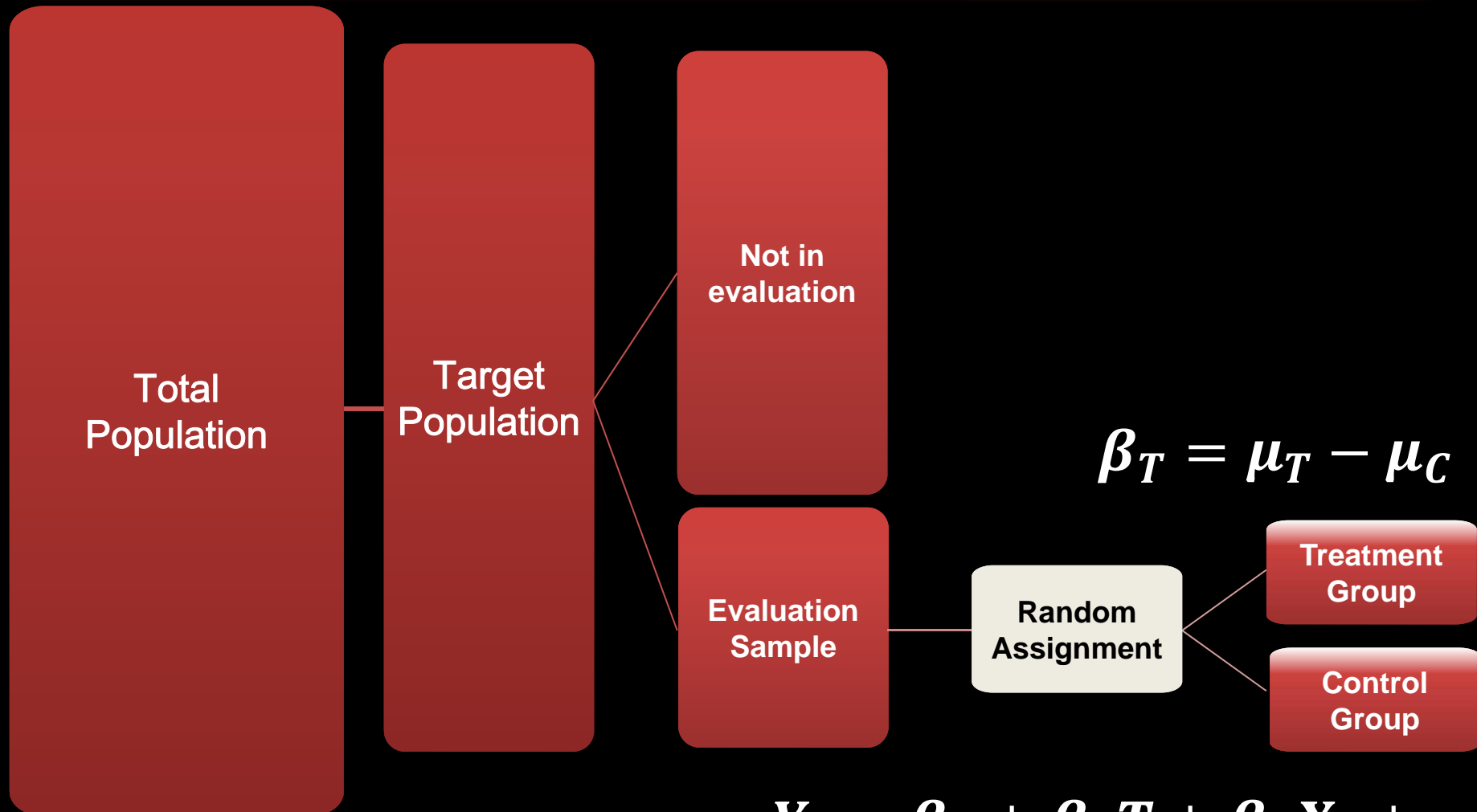
Randomly *sample* from area of interest

Randomly *assign* to **treatment** and **control**

Randomly *sample* from both treatment and control



Basic setup of a randomized evaluation



$$\beta_T = \mu_T - \mu_C$$

$$Y_i = \beta_0 + \beta_1 T + \beta_X X_i + \varepsilon$$

Impact of Balsakhi - Summary

Method	Impact Estimate
(1) Pre-post	26.42*
(2) Simple Difference	-5.05*
(3) Difference-in-Difference	6.82*
(4) Regression with controls	1.92
(5) Randomized Experiment	5.87*

*: Statistically significant at the 5% level

Impact of Balsakhi - Summary

Method	Impact Estimate
(1) Pre-post	26.42*
(2) Simple Difference	-5.05*
(3) Difference-in-Difference	6.82*
(4) Regression with controls	1.92
(5) Randomized Experiment	5.87*

*: Statistically significant at the 5% level

Bottom Line:

Which method we use matters

Be aware of the assumptions and possible biases

Conditions required

Method	Comparison	Works if....
Pre-Post	Program participants before program	The program was the only factor influencing any changes in the measured outcome over time
Simple Difference	Individuals who did not participate (data collected after program)	Non-participants are identical to participants except for program participation, and were equally likely to enter program before it started.
Differences in Differences	Same as above, plus: data collected before and after	If the program didn't exist, the two groups would have had identical trajectories over this period.
Multivariate Regression	Same as above plus: Also have additional "explanatory" variables	Omitted (because not measured or not observed) variables do not bias the results because they are either: uncorrelated with the outcome, or do not differ between participants and non-participants
Propensity Score Matching	Non-participants who have mix of characteristics which predict that they would be as likely to participate as participants	Same as above
Randomized Evaluation	Participants randomly assigned to control group	Randomization "works" – the two groups are statistically identical on observed and unobserved characteristics

Qualitative vs Quantitative

Qualitative methods

- Focus Group Discussions
- Case Studies
- Interviews – semi-structured, structured
- Participatory methods - Participatory Rural Appraisal (PRA), Rapid Rural Appraisal (RRA)
- Most Significant Change
- Observations

