ABDUL LATIF JAMEEL
Poverty Action Lab

TRANSLATING RESEARCH INTO ACTION

Impact Evaluation Methods: Why Randomize?

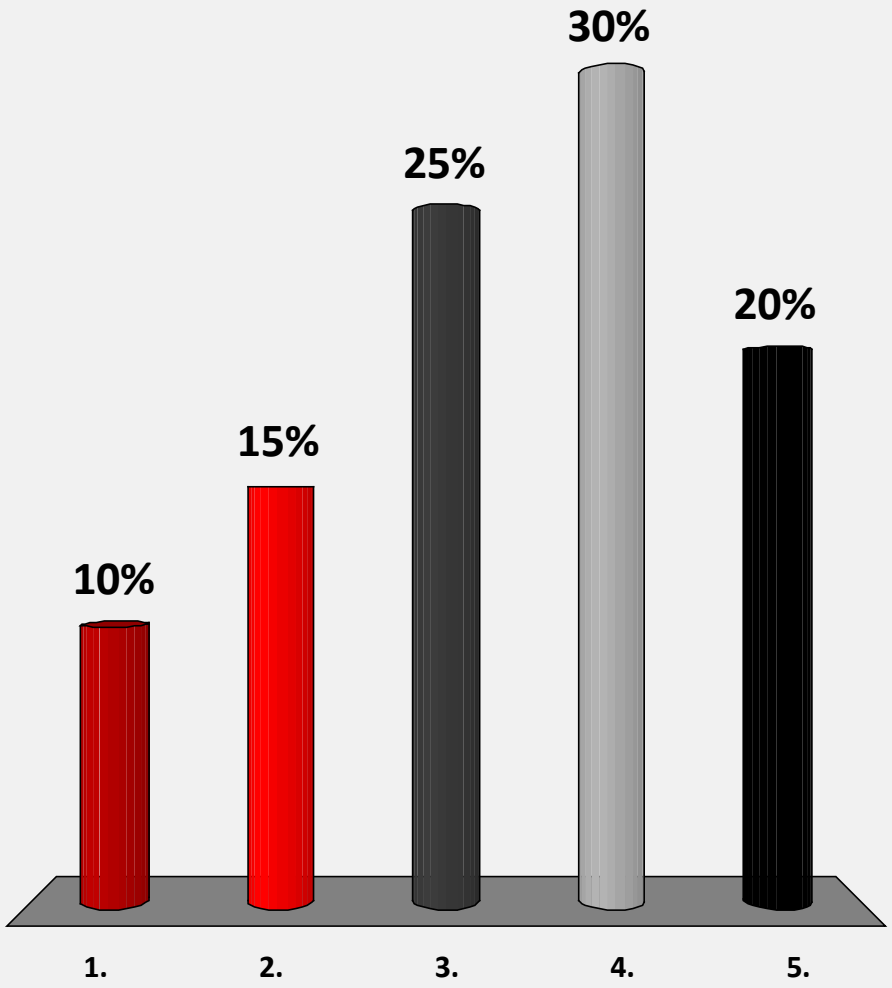
Meghan Mahoney
Policy Manager, J-PAL Global

Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. **Why Randomize?**
4. How to Randomize?
5. Project from Start to Finish
6. Generalizability

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

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



Session Overview

I. Background

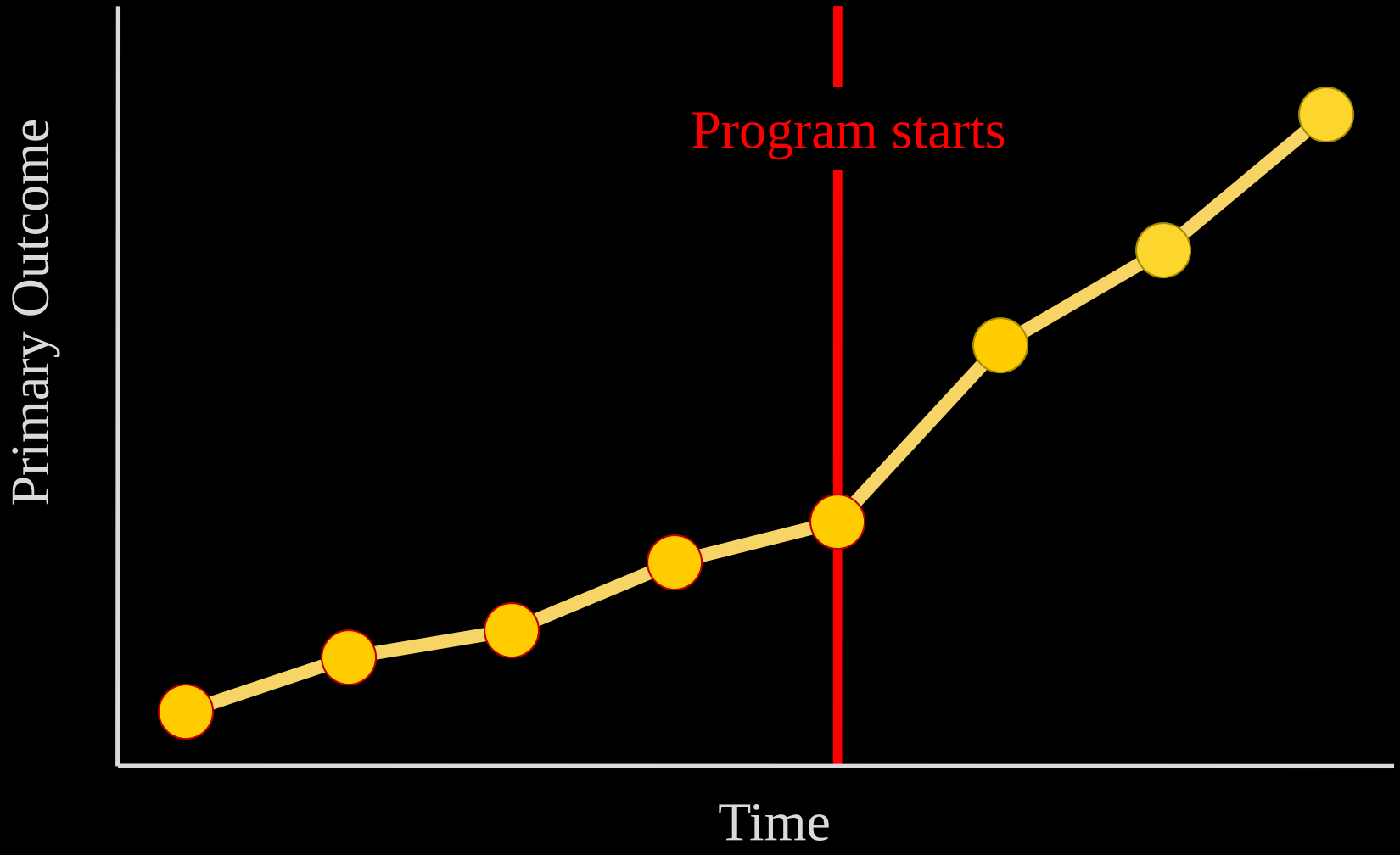
II. What is a randomized experiment?

III. Why randomize?

IV. Conclusions

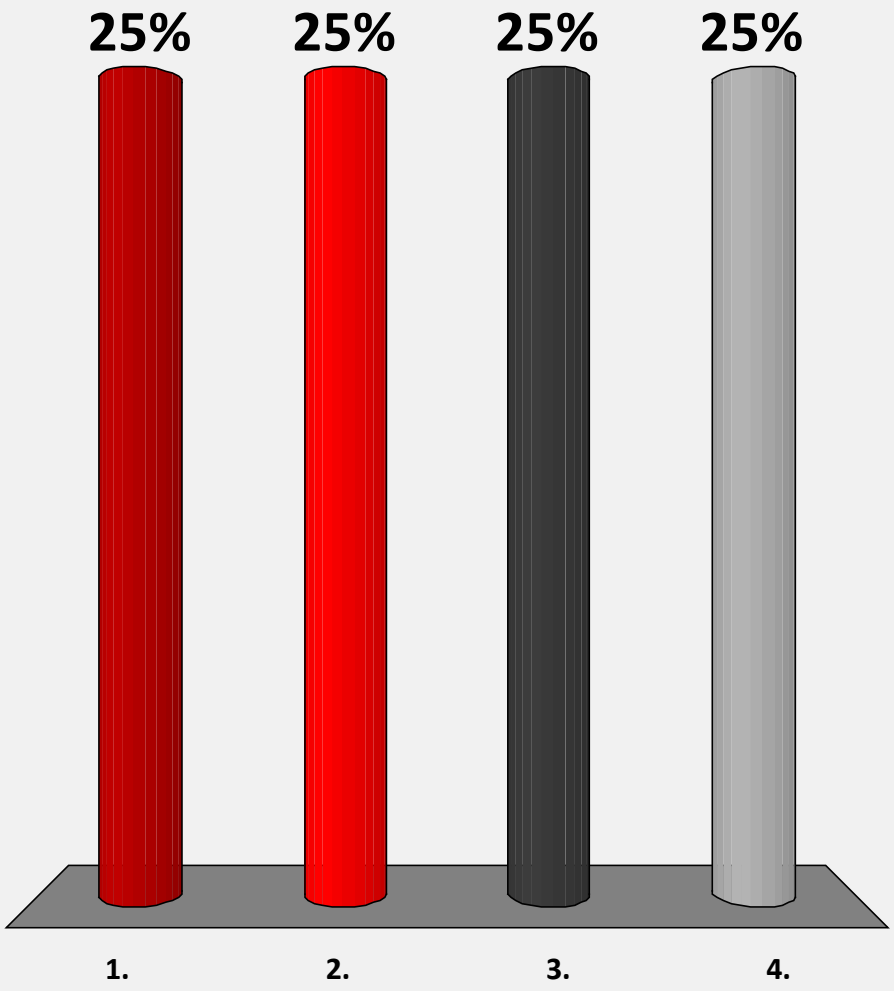
I - BACKGROUND

What is the impact of this program?



What is the impact of this program?

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



Read India

FIGURE 1

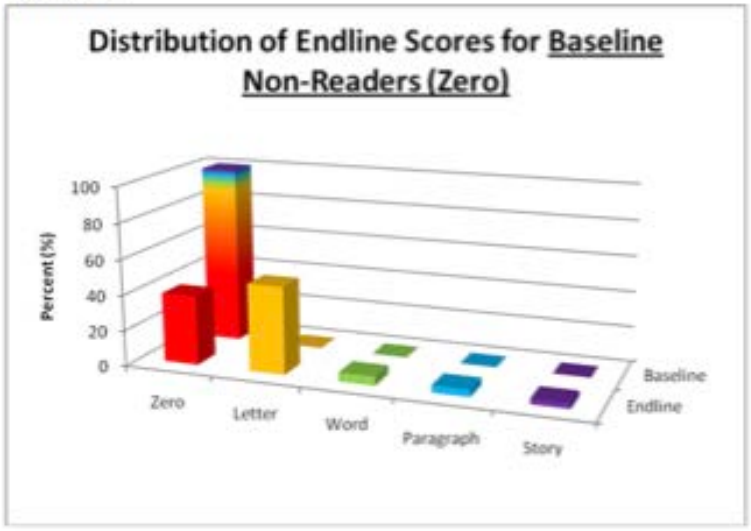
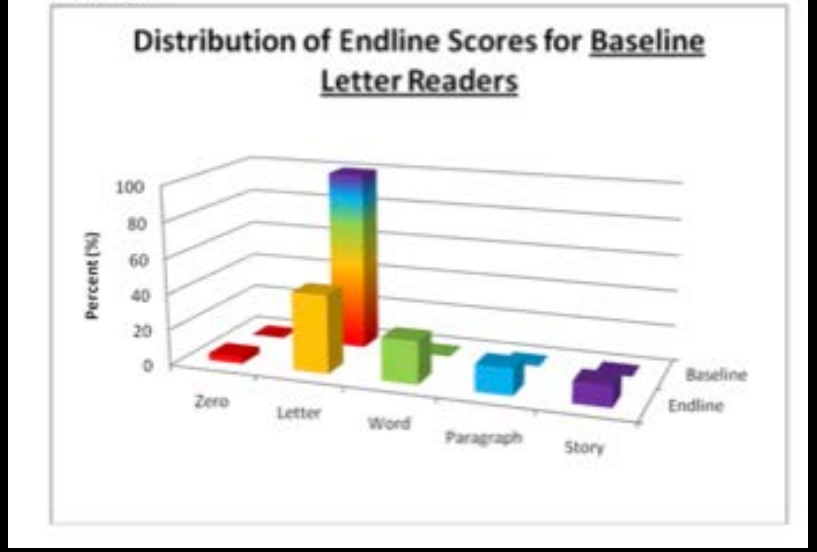


FIGURE 2



"Before vs. After" is rarely a good method for assessing impact.

How to measure impact?

Impact is defined as a comparison between:

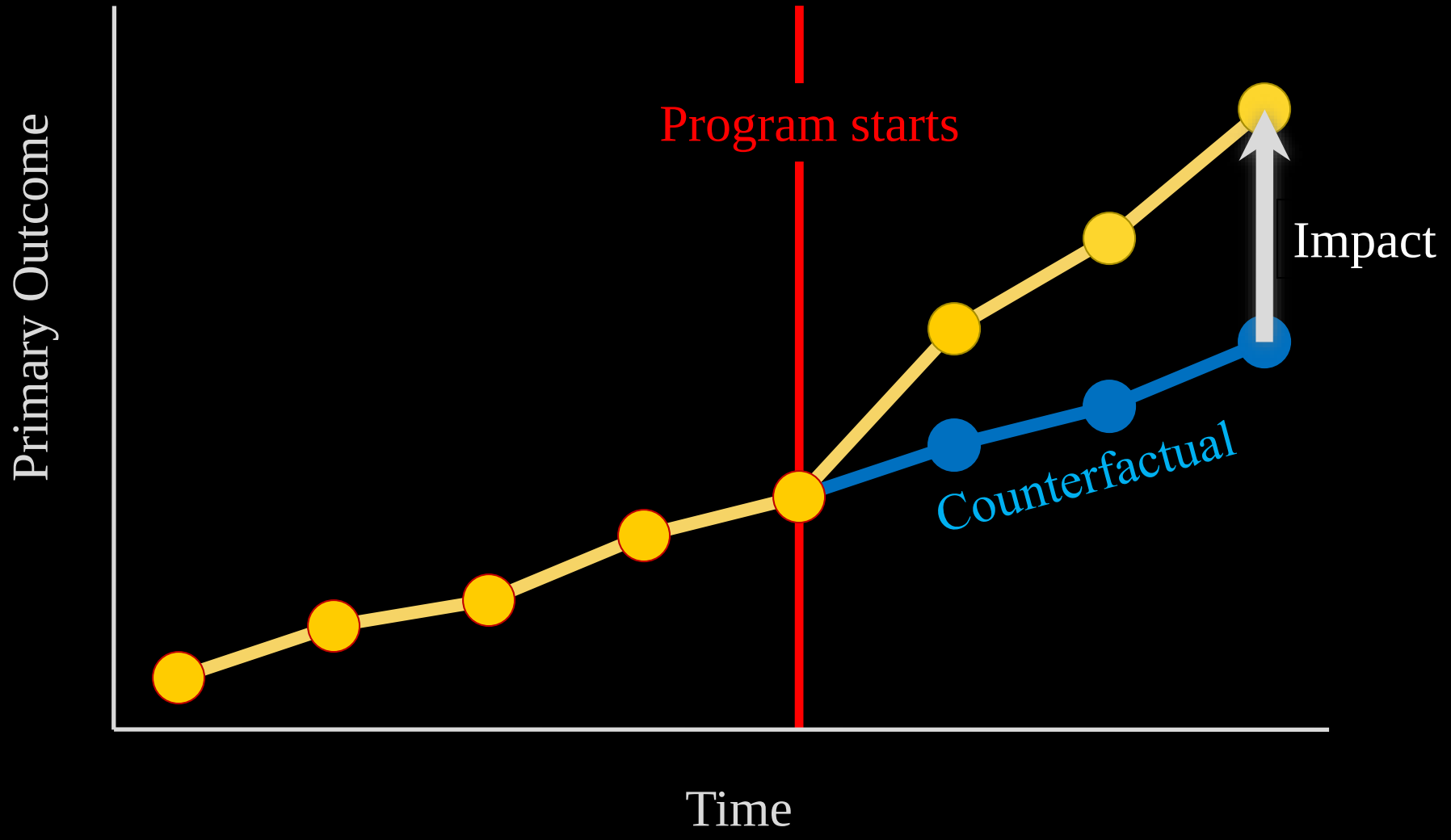
1. the outcome some time after the program has been introduced

2. the outcome at that same point in time had the program not been introduced (the "*counterfactual*")

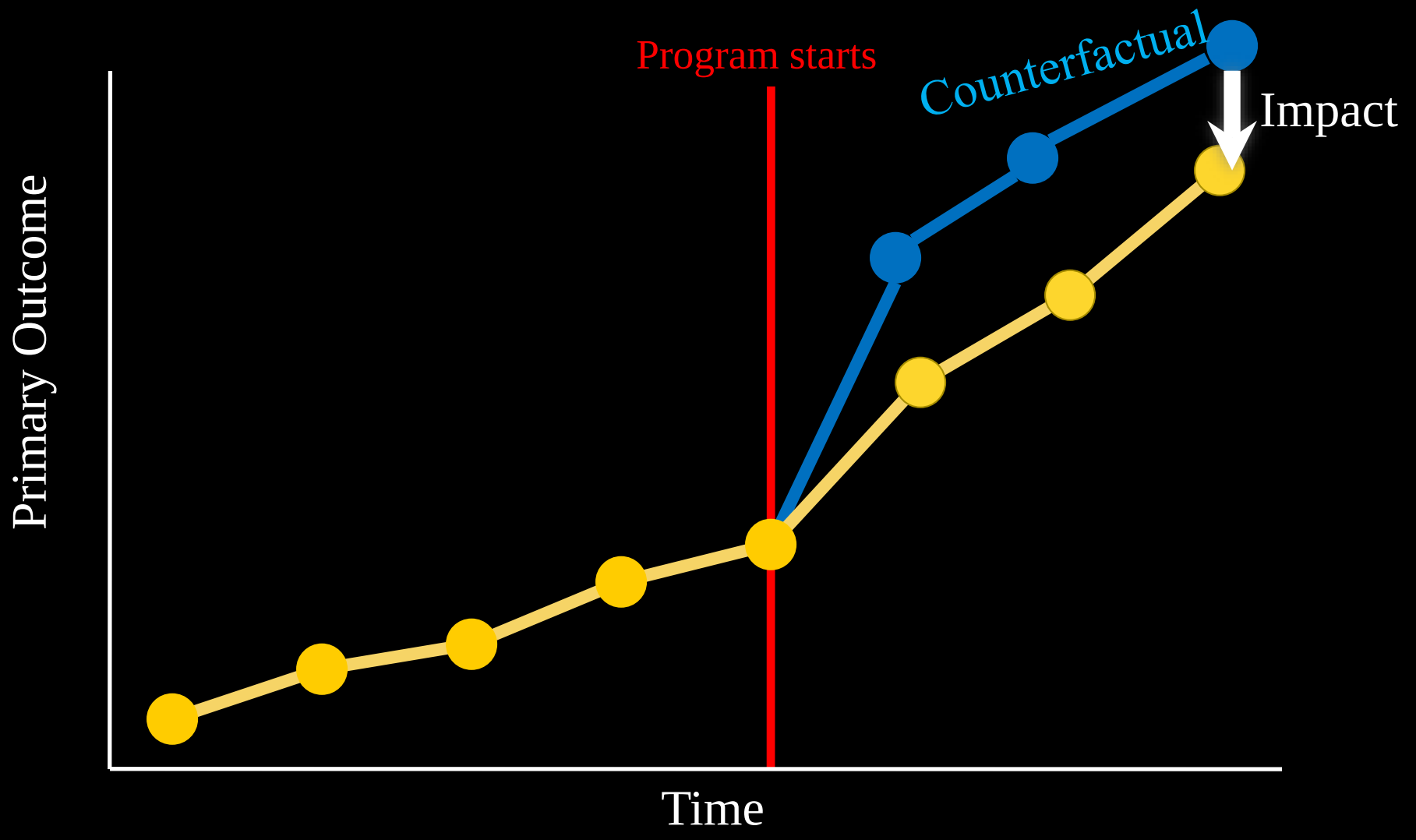
The counterfactual

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)

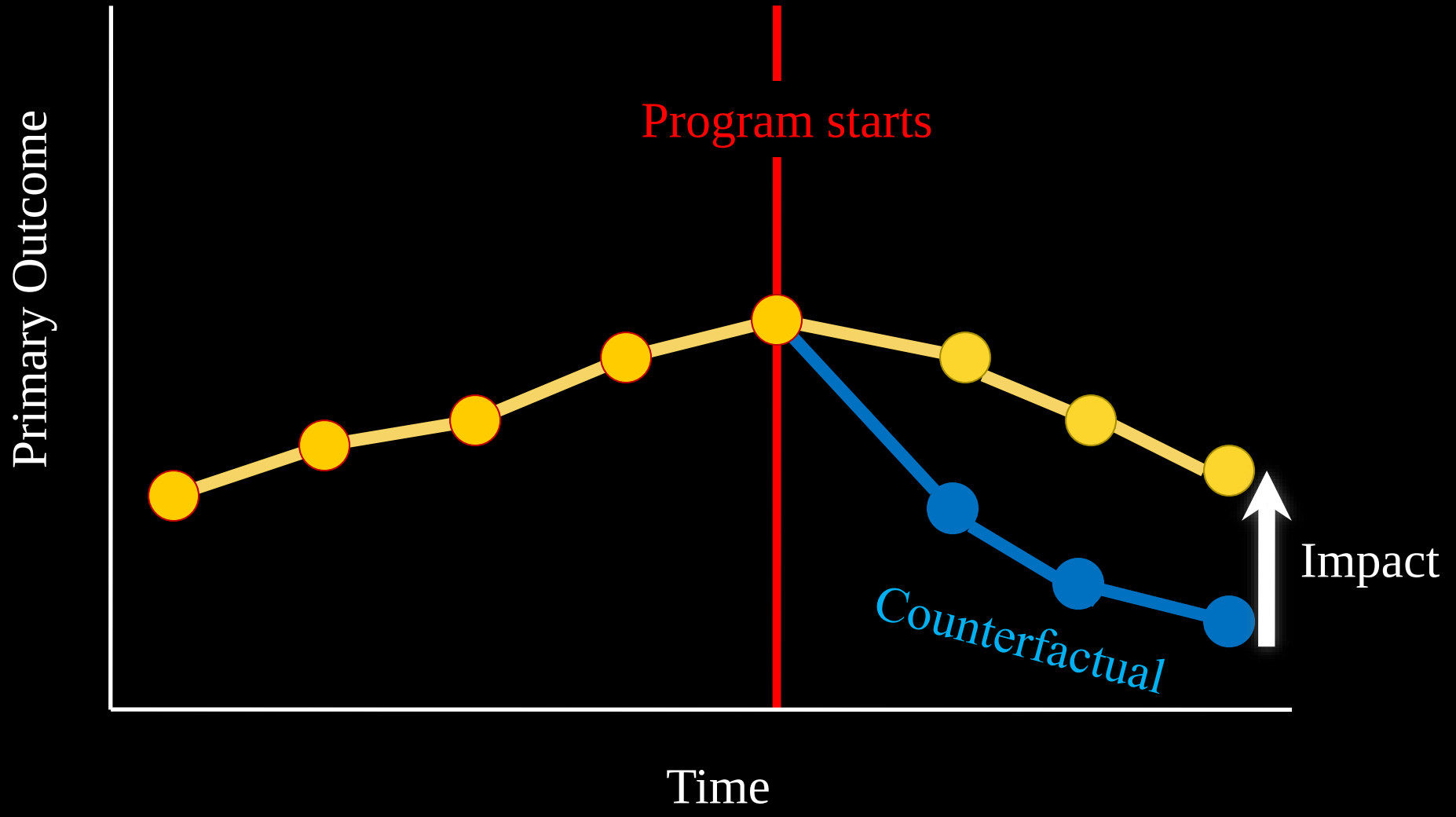
What is the impact of this program?



What is the impact of this program?



What is the impact of this program?



The problem with the counterfactual

- **In the real world, the c**ounterfactual cannot be observed
- *Solution*: We need to "mimic" or construct the counterfactual

Constructing the 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

Selecting the comparison group

- Select a group that is **exactly like** the group of participants in all ways except one: their exposure to the program being evaluated

- Goal: To be able to **attribute** differences in outcomes between the group of participants and the comparison group to the program (and not to other factors)

Impact evaluation methods

1. Randomized Experiments Also known as:
   - Random Assignment Studies
   - Randomized Field Trials
   - Social Experiments
   - Randomized Controlled Trials (RCTs)
   - Randomized Controlled Experiments

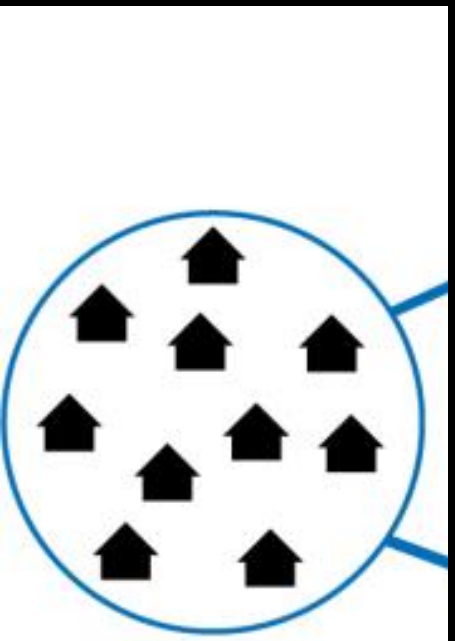
Impact evaluation methods

2. Non- or Quasi-Experimental Methods
    a. Pre-Post
    b. Simple Difference
    c. Differences-in-Differences
    d. Multivariate Regression
    e. Statistical Matching
    f. Interrupted Time Series
    g. Instrumental Variables
    h. Regression Discontinuity

II – WHAT IS A RANDOMIZED EXPERIMENT AND WHY IS IT SPECIAL?

The methodology of randomized evaluations

Before the program starts, eligible individuals are randomly assigned to two groups that are statistically identical before the program.



Key advantage of experiments

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.

Some variations on the basics

- Assigning to multiple treatment groups
- Assigning of units other than individuals or households
  - Health Centers
  - Schools
  - Local Governments
  - Villages

Key steps in conducting an experiment

1. Design the study carefully
2. Randomly assign people to treatment or control
3. Collect baseline data
4. Verify that assignment looks random
5. Monitor process so that integrity of experiment is not compromised

Key steps in conducting an experiment (cont.)

6. Collect follow-up data for both the treatment and control groups
7. Estimate program impacts by comparing mean outcomes of treatment group vs. mean outcomes of control group.
8. Assess whether program impacts are statistically significant and practically significant.

III – WHY RANDOMIZE?

EXAMPLE: IMPROVING STUDENT LEARNING IN INDIA

Example: Balsakhi Program

Balsakhi Program: Background

- Implemented by Pratham, an NGO from India
- Program provided tutors (Balsakhi) to help at-risk children with school work
- In Vadodara, the balsakhi program was run in government primary schools in 2002-2003
- Teachers decided which children would get the balsakhi

Balsakhi: Outcomes

- Children were tested at the beginning of the school year (Pretest) and at the end of the year (Post-test)

- QUESTION: How can we estimate the impact of the balsakhi program on test scores?

Methods to estimate impacts

- Let's look at different ways of estimating the impacts using the data from the schools that got a balsakhi
    1. Pre – Post (Before vs. After)
    2. Simple difference
    3. Difference-in-difference
    4. Other non-experimental methods
    5. Randomized Experiment

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)

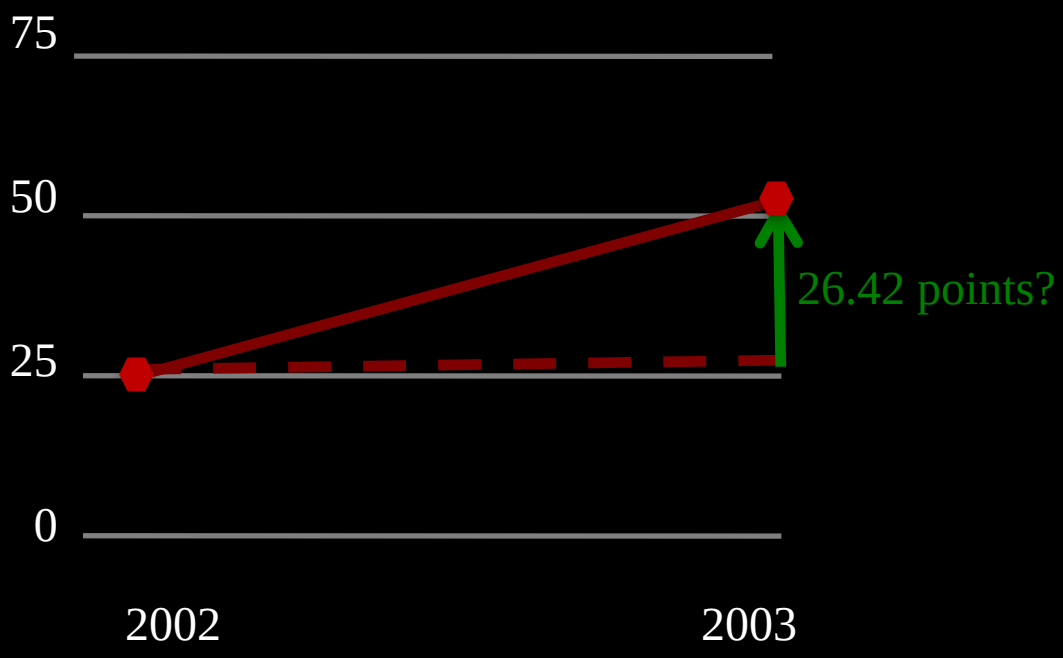
Average **post-test** score for children with a balsakhi	51.22
Average **pretest** score for children with a balsakhi	24.80
Difference	**26.42**

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

What would have happened without balsakhi?

Method 1: Before vs. After

Impact = 26.42 points?



2 - Simple difference

Compare test scores of…

Children who got balsakhi

With test scores of…

Children who did not get balsakhi

2 - Simple difference

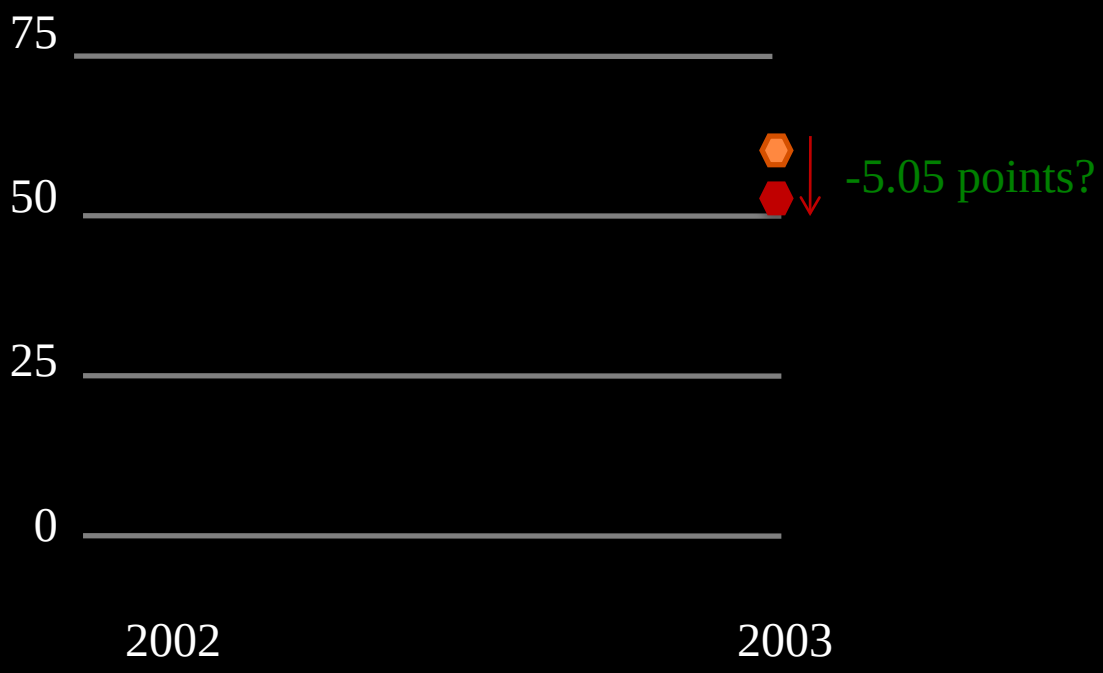
Average score for children with a balsakhi	**51.22**
Average score for children without a balsakhi	56.27
Difference	-5.05

- Question: Under what conditions can this difference (-5.05) be interpreted as the impact of the balsakhi program?

What would have happened without balsakhi?

Method 2: Simple Comparison

Impact = -5.05 points?



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

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

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

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**

4 – Other Methods

- There are more sophisticated non-experimental methods to estimate program impacts:
  - Regression
  - Matching
  - Instrumental Variables
  - Regression Discontinuity

- These methods rely on being able to "mimic" the counterfactual under certain assumptions
- Problem: Assumptions are not testable

4: Multivariate Regression

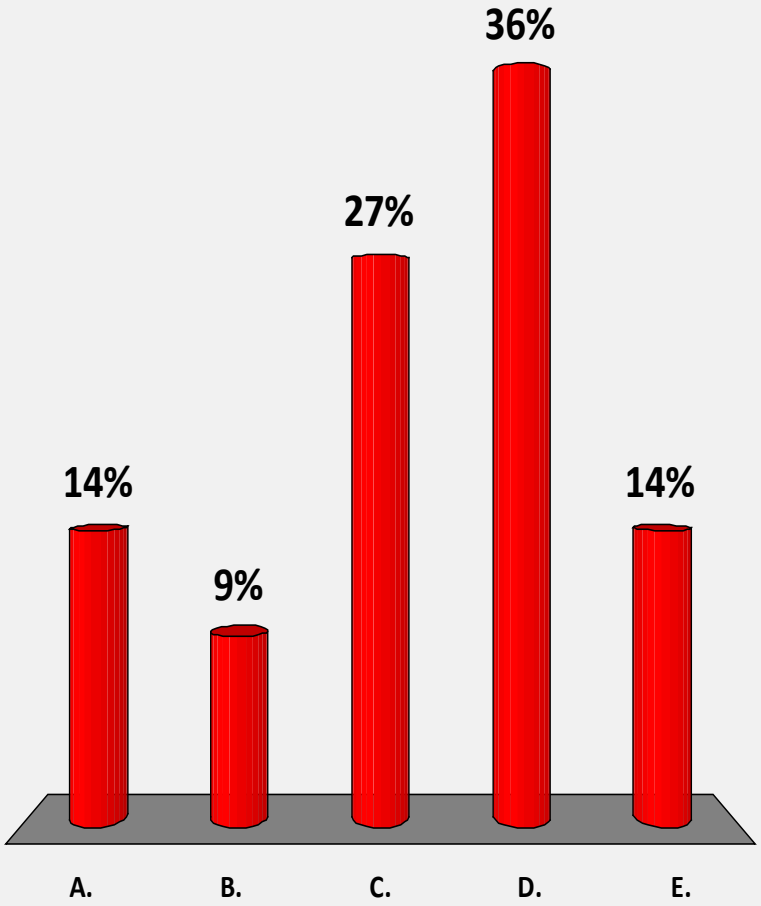
- Compares program participants to non-participants, while "controlling" for a variety of observable characteristics that could influence outcome
- But… what about unobservable characteristics?
- Run a regression

Which of these methods do you think is closest to the truth?

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

*: Statistically significant at the 5% level

A. Pre-Post
B. Simple Difference
C. Difference-in-Differences
D. Regression
E. Don't know



5 – Randomized Experiment

- Suppose we evaluated the Balsakhi program using a randomized experiment
- Question #1: What would this entail? How would we do it?
- Question #2: What would be the advantage of using this method to evaluate the impact of the balsakhi program?

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	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	1.92
(5)Randomized Experiment	5.87*

*: Statistically significant at the 5% level

Bottom Line: Which method we use matters!

Why randomize? – Conceptual Argument

Since groups are statistically equal, randomized evaluations require fewer assumptions than other impact evaluation methodologies.

If properly designed and conducted, randomized experiments provide the most credible method to estimate the impact of a program.

Why "most credible"?

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.

IV – CONCLUSIONS

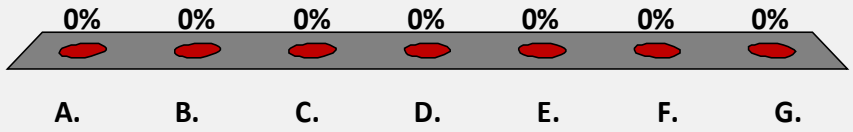
Conclusions - Why Randomize?

- There are many ways to estimate a program's impact
- This course argues in favor of one: randomized experiments
  - Conceptual argument: If properly designed and conducted, randomized experiments provide the most credible method to estimate the impact of a program
  - Empirical argument: Different methods can generate different impact estimates

THANK YOU!
QUESTIONS?

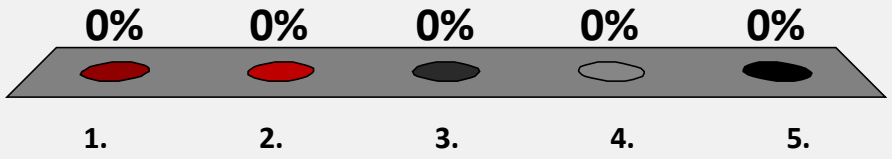
What is the most convincing argument you have heard **against** RCTs? Enter your **top 3 choices**.

A. Too expensive
B. Takes too long
C. Not ethical
D. Too difficult to design/implement
E. Not externally valid (Not generalizable)
F. Less practical to implement than other methods and not much better
G. Can tell us *what the impact is* impact, but not *why* or *how* it occurred (i.e. it is a black box)



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

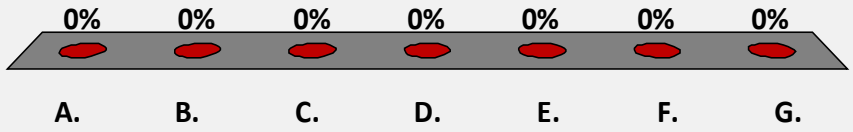


Why randomize? - Backup Slides

Dan Levy
Harvard Kennedy School

What is the most convincing argument you have heard **against** RCTs? Enter your **top 3 choices**.

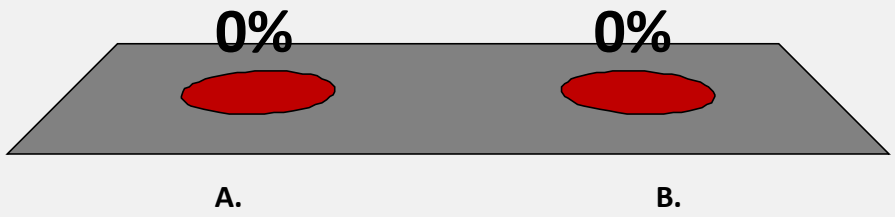
A. Too expensive
B. Takes too long
C. Not ethical
D. Too difficult to design/implement
E. Not externally valid (Not generalizable)
F. Less practical to implement than other methods and not much better
G. Can tell us *what the impact is* impact, but not *why* or *how* it occurred (i.e. it is a black box)



What do you want to do?

A. Example

B. Objections to RCTs



Example #2 - Pratham's Read India program

Example #2 - Pratham's Read India program

Method	Impact
(1) Pre-Post	0.60*
(2) Simple Difference	-0.90*
(3) Difference-in-Differences	0.31*
(4) Regression	0.06
(5) Randomized Experiment	

*: Statistically significant at the 5% level

Example #1 - Pratham's Read India program

Method	Impact
(1) Pre-Post	0.60*
(2) Simple Difference	-0.90*
(3) Difference-in-Differences	0.31*
(4) Regression	0.06
(5) Randomized Experiment	0.88*

*: Statistically significant at the 5% level

Example - Pratham's Read India program

Method	Impact
(1) Pre-Post	0.60*
(2) Simple Difference	-0.90*
(3) Difference-in-Differences	0.31*
(4) Regression	0.06
(5) Randomized Experiment	0.88*

Bottom Line: Which method we use matters!

Example #2: A voting campaign in the USA

A voting campaign in the USA

Method	Impact (vote %)
(1) Pre-post	-7.2 pp
(2) Simple difference	10.8 pp *
(3) Difference-in-differences	3.8 pp*
(4) Multiple regression	6.1 pp *
(5) Matching	2.8 pp *
(5) Randomized Experiment	

A voting campaign in the USA

Method	Impact (vote %)
(1) Pre-post	-7.2 pp
(2) Simple difference	10.8 pp *
(3) Difference-in-differences	3.8 pp*
(4) Multiple regression	6.1 pp *
(5) Matching	2.8 pp *
(5) Randomized Experiment	0.4 pp

A voting campaign in the USA

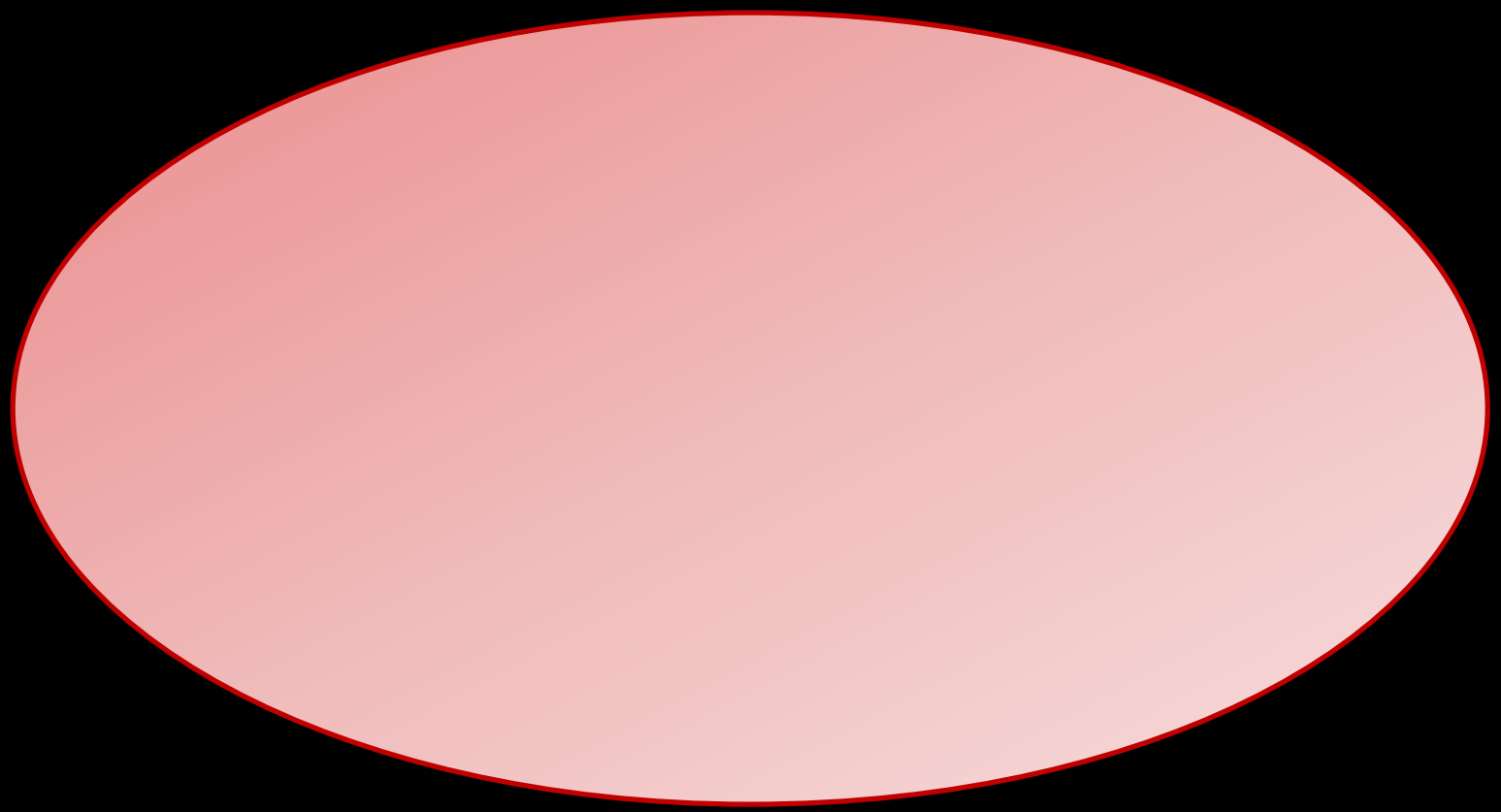
Method	Impact (vote %)
(1) Pre-post	-7.2 pp
(2) Simple difference	10.8 pp *
(3) Difference-in-differences	3.8 pp*
(4) Multiple regression	6.1 pp *
(5) Matching	2.8 pp *
(5) Randomized Experiment	0.4 pp

Bottom Line: Which method we use matters!

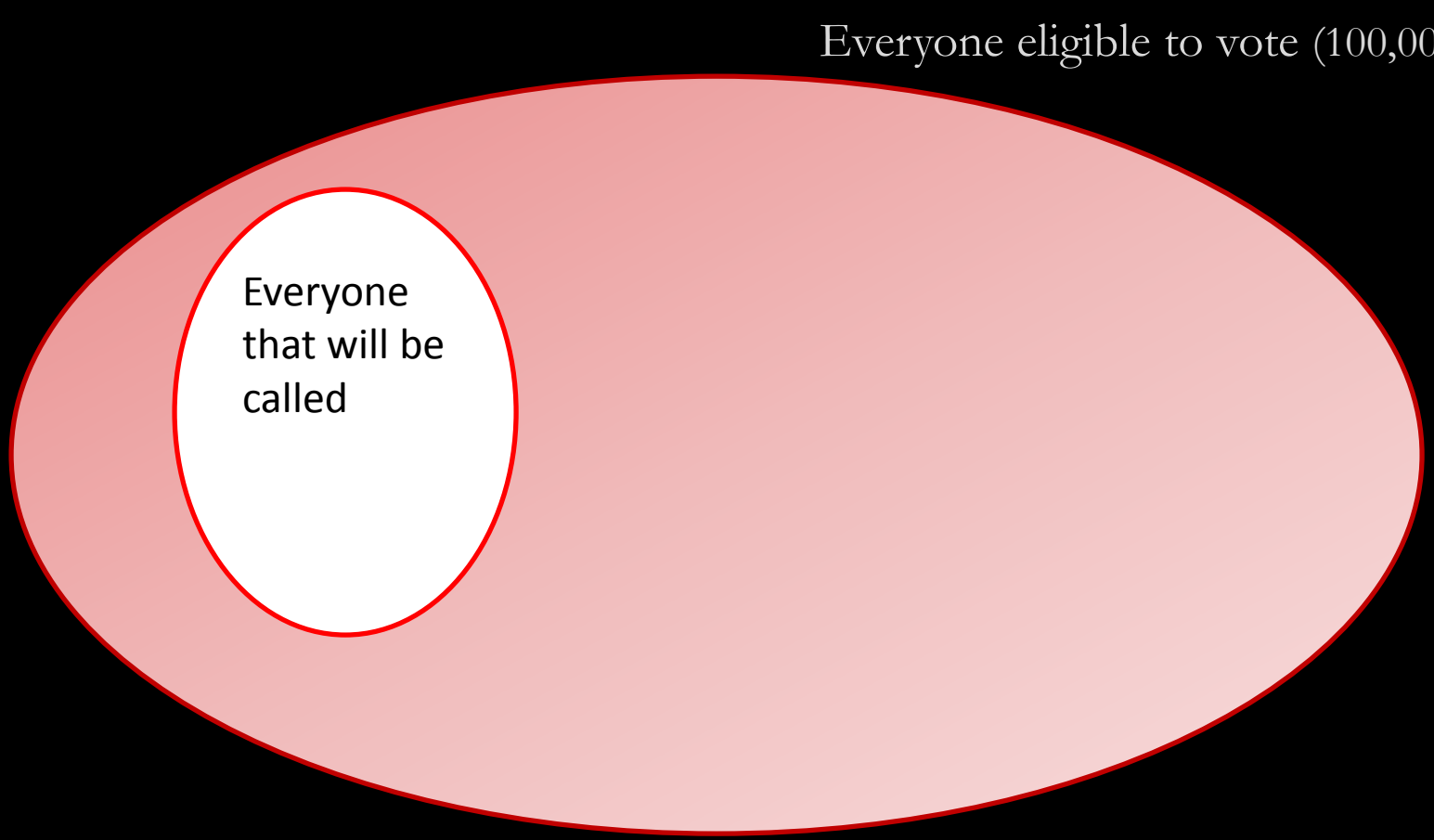
Program "Get Out the Vote"

- Low voter turnout is seen as a problem in many countries in the world
- Some countries have looked for ways to increase voter turnout
- "Get Out the Vote" Program
  - Compiled a list of all the 100,000 individuals who could vote in an election
  - Call a sample individuals in this list
  - In this phone call, responder is encouraged to vote

Program “Get Out the Vote”



Program "Get Out the Vote"



Program "Get Out the Vote"(Cont)

- Key question: What is the **impact** of the "Get Out the Vote" program on the voter turnout rate?

- Methodological Question: How should we estimate the impact of the program?

Resources available for the evaluation

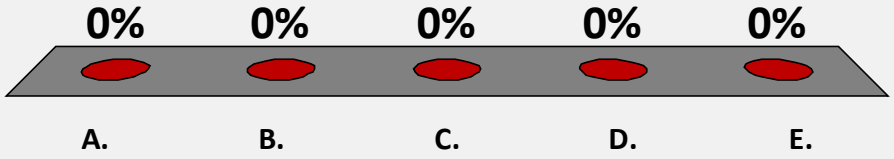
- List of all the persons eligible to vote with information on:
  - Income
  - Education
  - Sex
  - Age
  - Whether person voted in the last election

- Money to make up to 8,000 calls that could be used to:
  - Implement the program (i.e. call before the election encouraging person to vote)
  - Collect data (i.e. call people after the election to ask whether they voted or not)

- List of 2,000 people who came to a political rally one month before the election
  - You already called them and encouraged them to vote
  - These calls count as part of your 8,000 calls

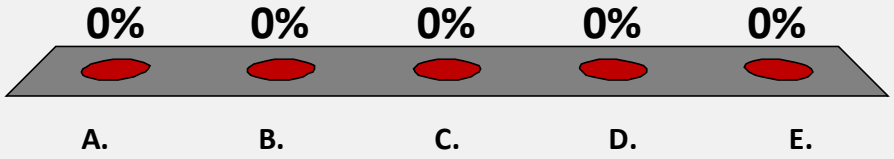
Which design would you choose?

A. Design 1
B. Design 2
C. Design 3
D. Design 4
E. Design 5



Which design would you choose?

A. Design 1
B. Design 2
C. Design 3
D. Design 4
E. Design 5



Example #2 - Pratham's Read India program

Example #2 - Pratham's Read India program

Method	Impact
(1) Pre-Post	0.60*
(2) Simple Difference	-0.90*
(3) Difference-in-Differences	0.31*
(4) Regression	0.06
(5) Randomized Experiment	

*: Statistically significant at the 5% level

Example #2 - Pratham's Read India program

Method	Impact
(1) Pre-Post	0.60*
(2) Simple Difference	-0.90*
(3) Difference-in-Differences	0.31*
(4) Regression	0.06
(5) Randomized Experiment	0.88*

*: Statistically significant at the 5% level

Example #3: A voting campaign in the USA

A voting campaign in the USA

Method	Impact (vote %)
(1) Pre-post	-7.2 pp
(2) Simple difference	10.8 pp *
(3) Difference-in-differences	3.8 pp*
(4) Multiple regression	6.1 pp *
(5) Matching	2.8 pp *
(5) Randomized Experiment	

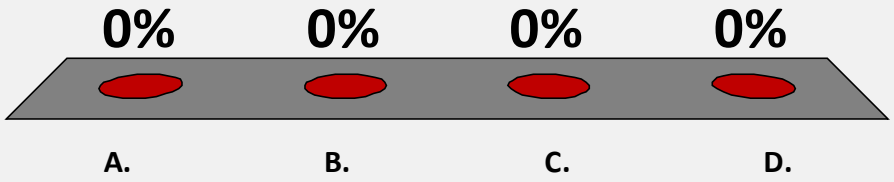
A voting campaign in the USA

Method	Impact (vote %)
(1) Pre-post	-7.2 pp
(2) Simple difference	10.8 pp *
(3) Difference-in-differences	3.8 pp*
(4) Multiple regression	6.1 pp *
(5) Matching	2.8 pp *
(5) Randomized Experiment	0.4 pp

THANK YOU!

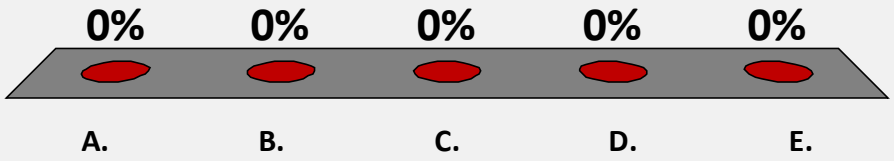
What is the impact of this program?

A. Positive
B. Negative
C. Zero
D. Not enough info

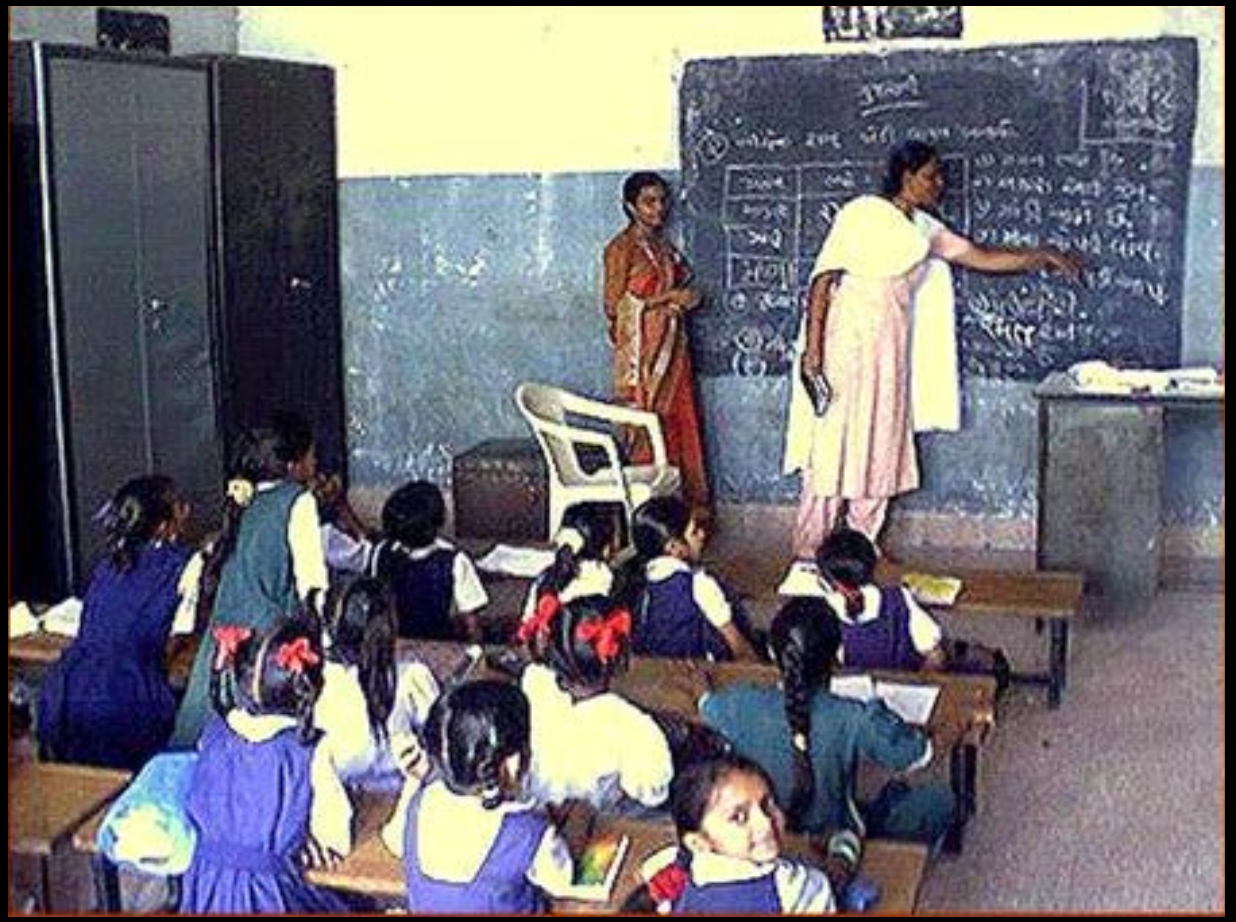


What is the impact of this program?

A. Positive
B. Negative
C. Zero
D. I don't know
E. Who knows?



Example #1: Balsakhi Program



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	1.92
(5)Randomized Experiment	5.87*

*: Statistically significant at the 5% level