



Why Randomize?

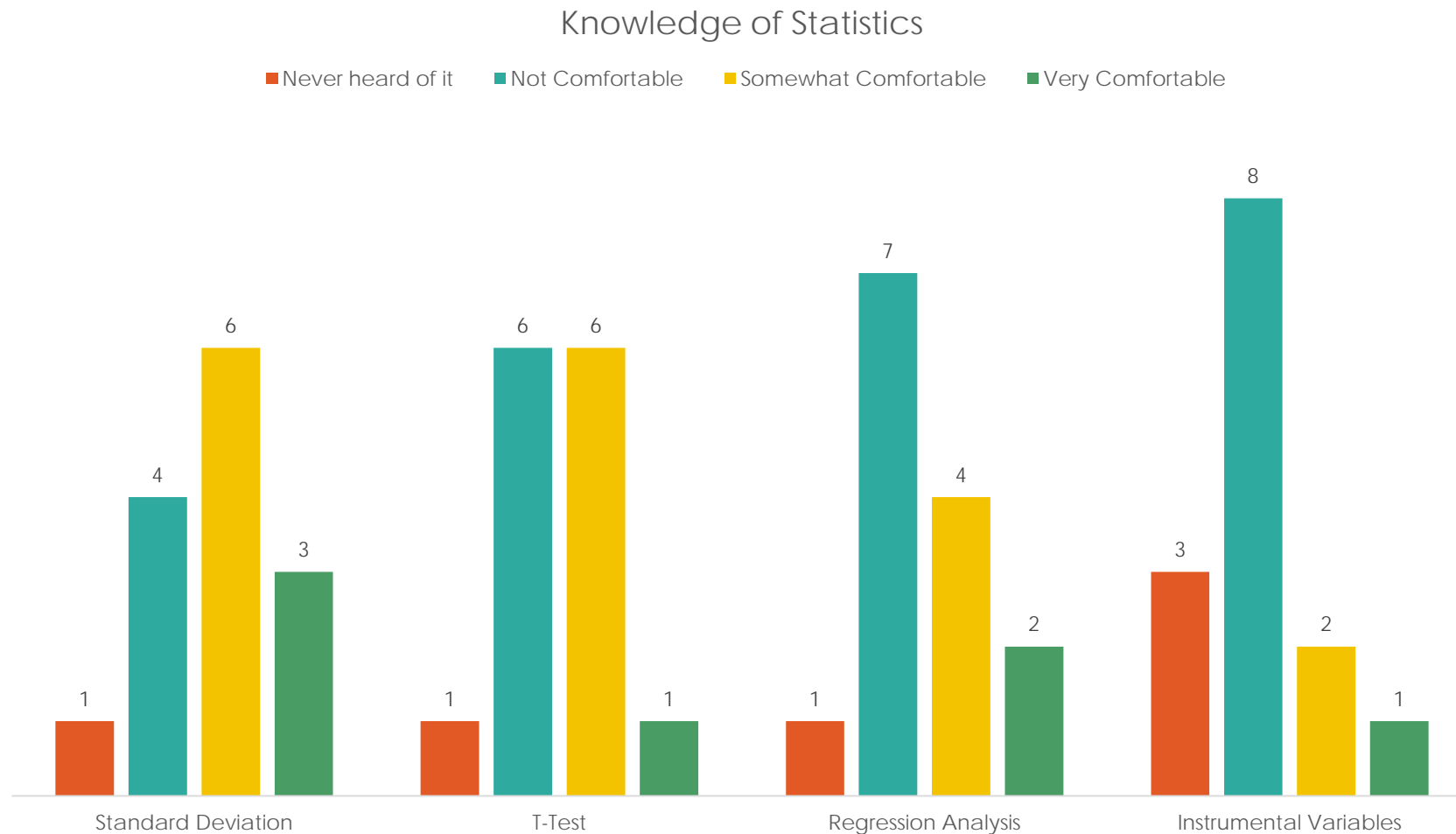
Thomas Chupein
Policy Manager
J-PAL Global at MIT



Course Overview

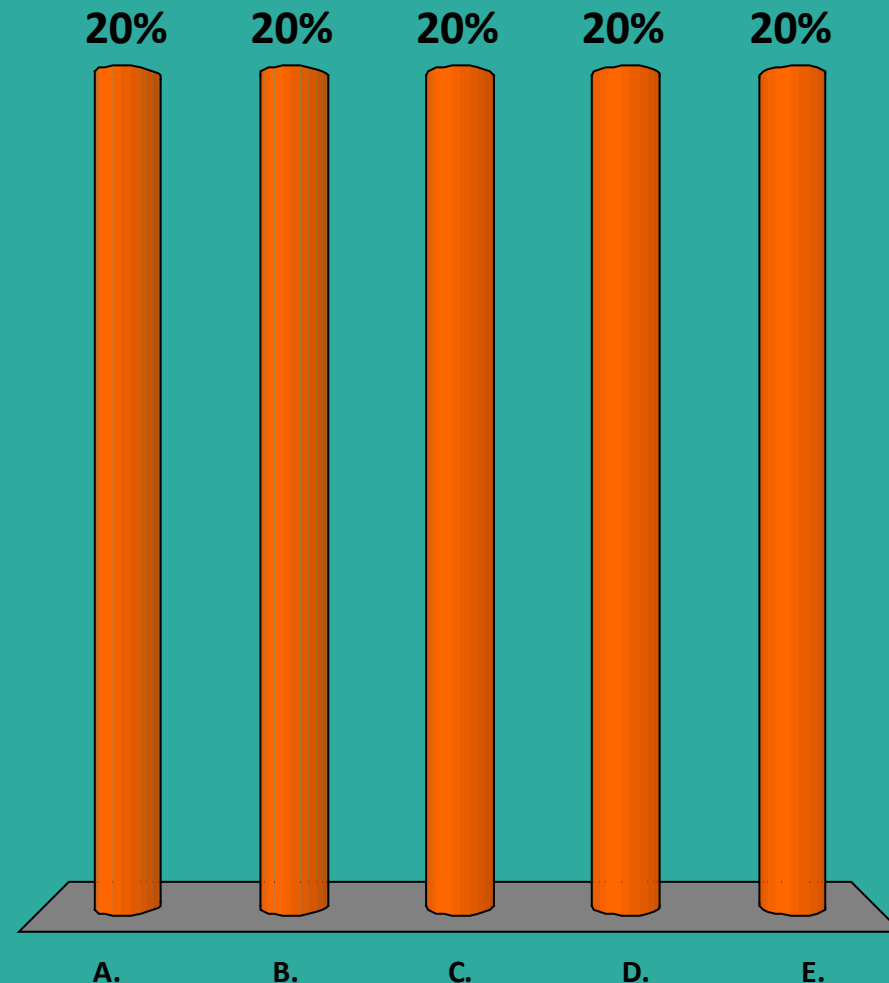
1. Introducing Randomized Impact Evaluations (Thomas Chupein)
2. Outcomes, Impact, and Indicators (Rohit Nainpally)
- 3. Why Randomize? (Thomas Chupein)**
4. How to Randomize (Rohit Nainpally)
5. Sampling and Sample Size (Rohit Nainpally)
6. Threats and Analysis (Lina Marliani)
7. Evaluation from Start to Finish (Lina Marliani)
8. Evidence from Community-Driven Development, Health, and Education Programs (Thomas Chupein and John Floretta)
9. Using Evidence from Randomized Evaluations for Decision-Making and Policy Change (John Floretta)
10. Discussion: Where do RCTs fit in a Good M&E Strategy? (All)

How comfortable are you with the following statistical concepts/methods:



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

- A. Strongly Disagree
- B. Disagree
- C. Neutral
- D. Agree
- E. Strongly Agree



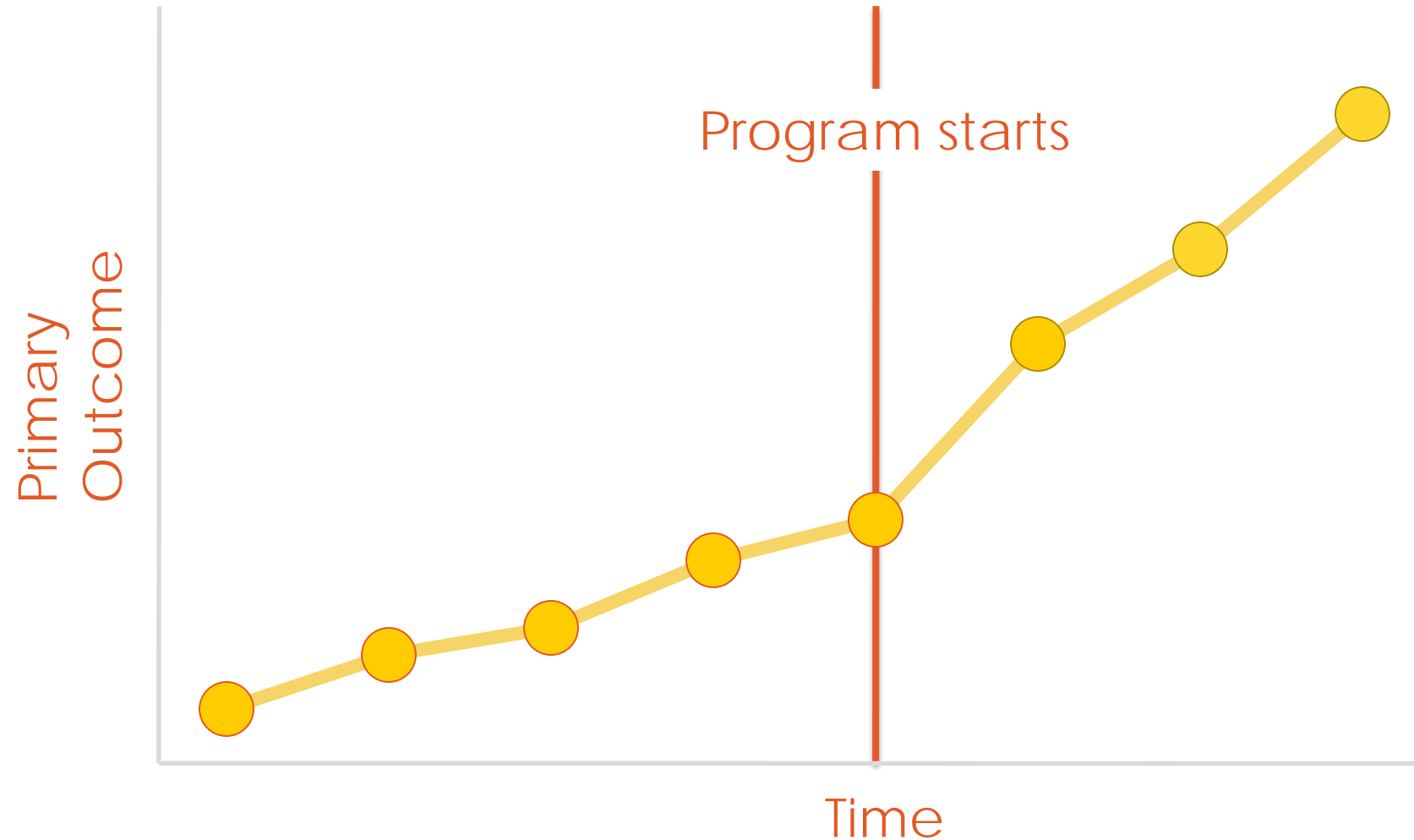
Session Overview

- Background
- What is a randomized experiment?
- Why randomize?
- Conclusions

1. Background

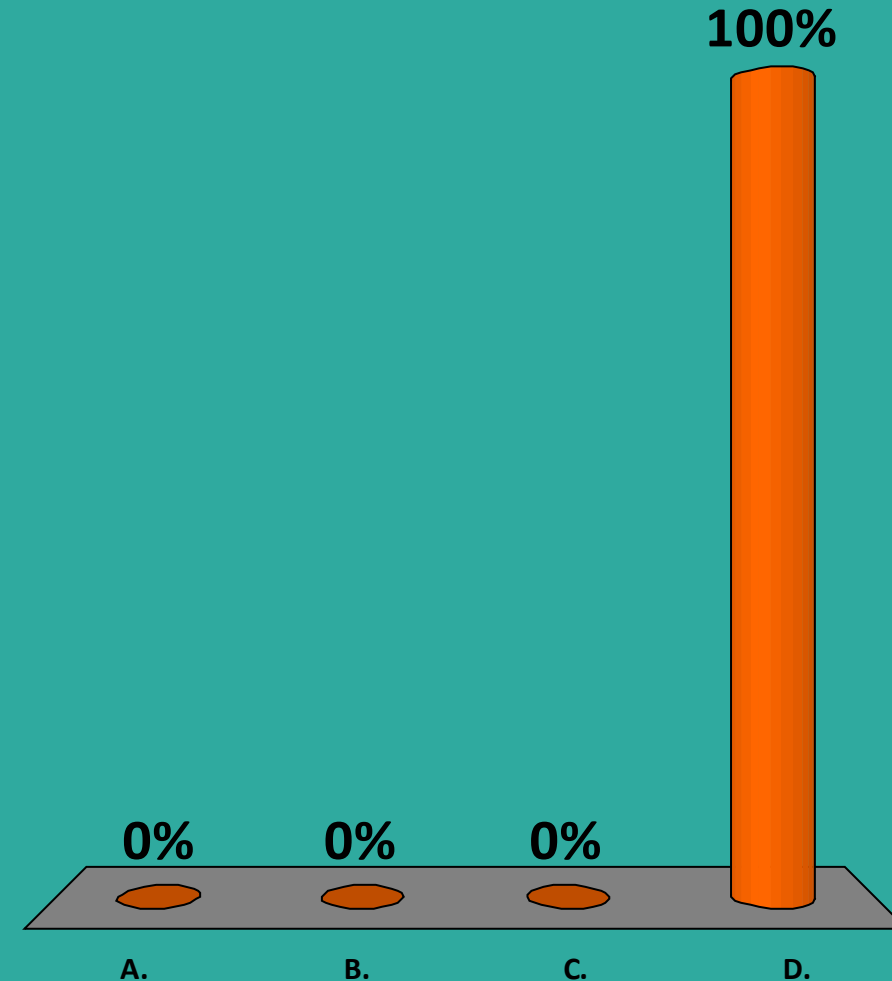


What is the impact of this program?



What is the impact of this program?

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



Read India

FIGURE 1

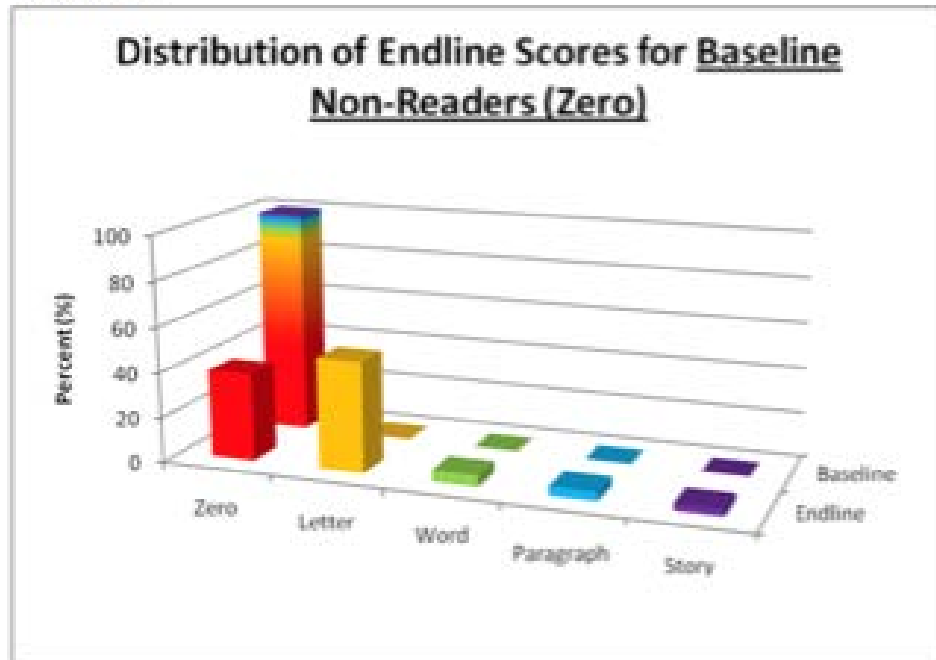
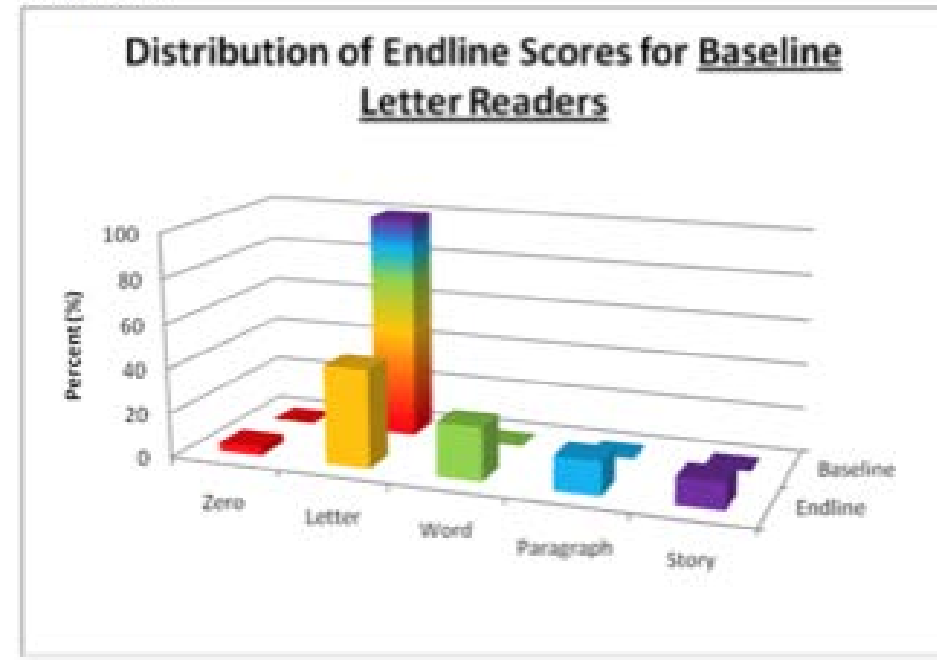
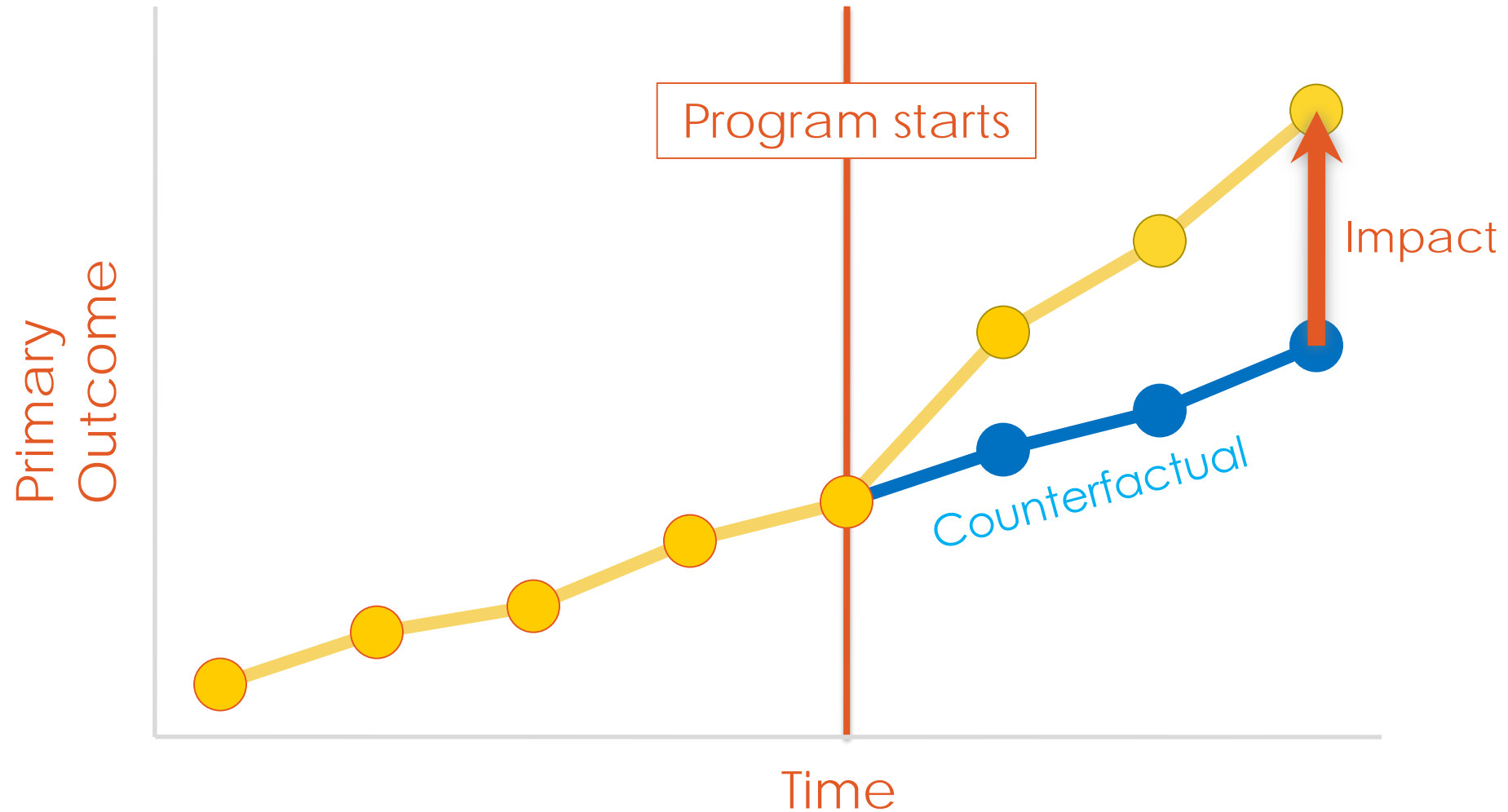


FIGURE 2



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

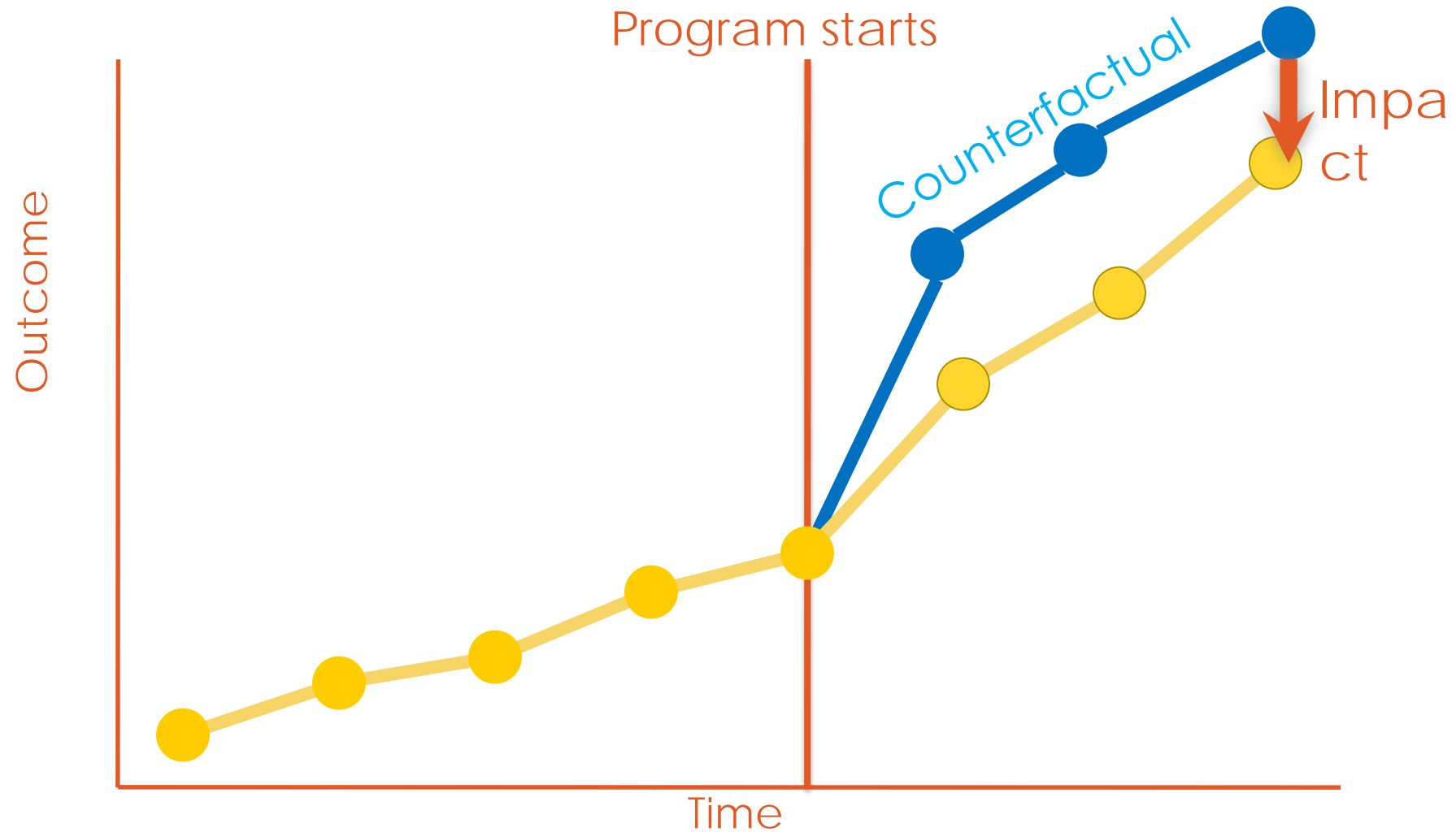
What is the impact of this program?



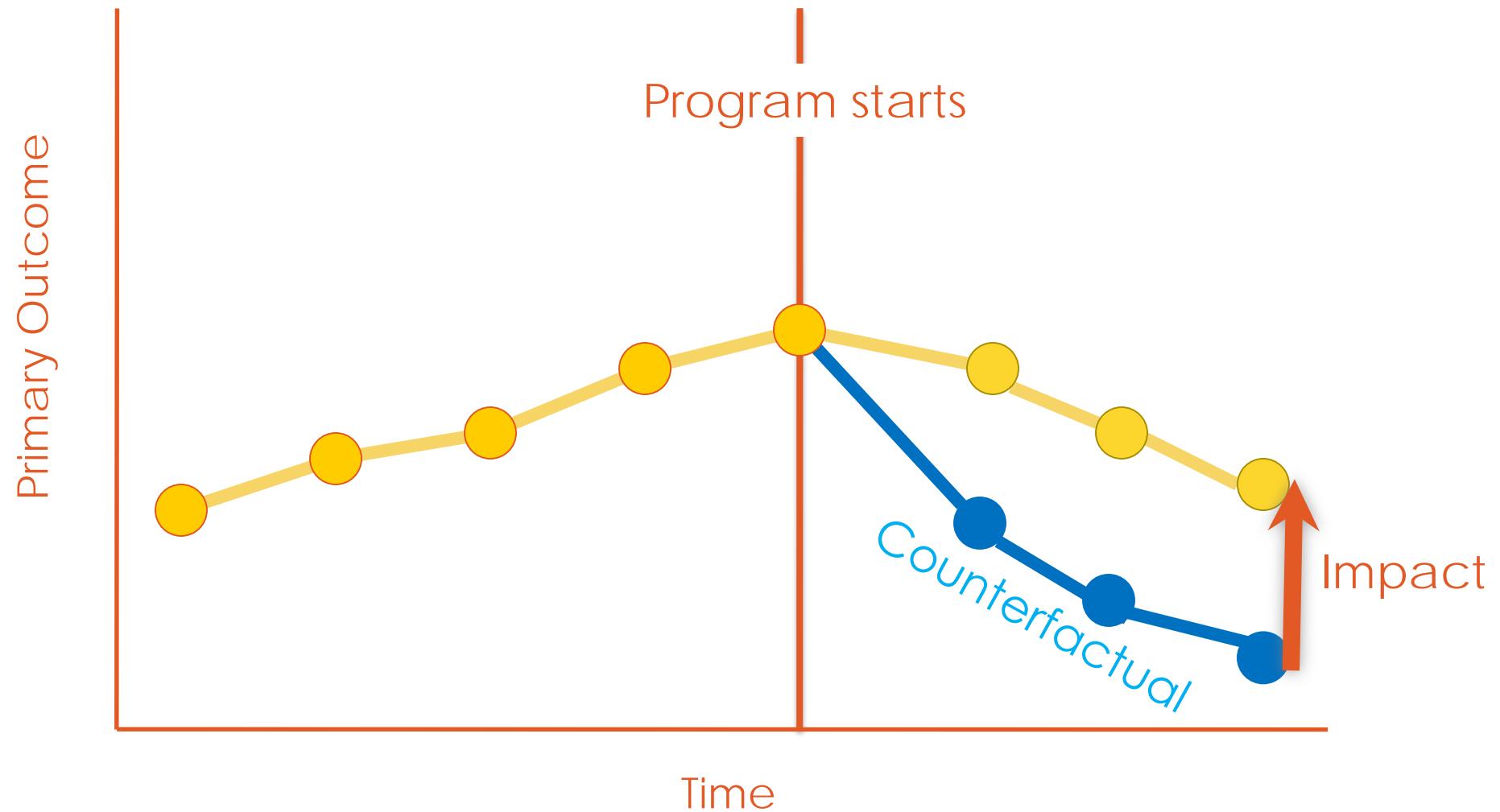
How to measure impact?

- Impact is defined as a comparison between:
 - the outcome some time after the program has been introduced
 - the outcome at that same point in time had the program not been introduced (the “counterfactual”)

Impact: What is it?



Impact: What is it?



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 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

- Idea: 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

- Pre-Post
- Simple Difference
- Difference-in-Differences
- Multivariate Regression
- Statistical Matching
- Interrupted Time Series
- Instrumental Variables
- Regression Discontinuity

2. What is a Randomized Experiment?



The basics

- Start with simple case:
- Take a sample of program applicants
- Randomly assign them to either:
 - Treatment Group – is offered treatment
 - Control Group – not allowed to receive treatment (during the evaluation period)

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.

Evaluation of “Women as Policymakers”: Treatment vs. Control villages at baseline

Variables	Treatment Group	Control Group	Difference
Female Literacy Rate	0.35	0.34	0.01 (0.01)
Number of Public Health Facilities	0.06	0.08	-0.02 (0.02)
Tap Water	0.05	0.03	0.02 (0.02)
Number of Primary Schools	0.95	0.91	0.04 (0.08)
Number of High Schools	0.09	0.10	-0.01 (0.02)

Standard Errors in parentheses. Statistics displayed for West Bengal

*/**/***: Statistically significant at the 10% / 5% / 1% level

Source: Chattopadhyay and Duflo (2004)

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 experiments is not compromised
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 the control group
8. Assess whether program impacts are statistically significant and practically significant

3. Why Randomize?



Why Randomize?- Conceptual Argument

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.

Example 1: 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 (pre-test) 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



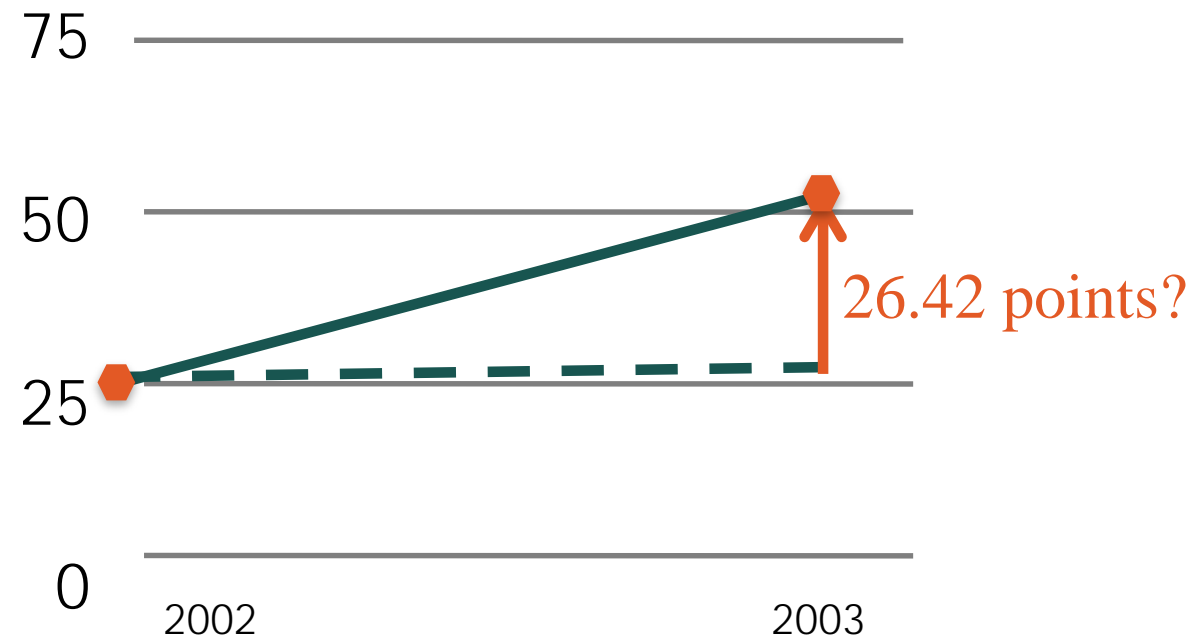
1 - Pre-post (Before vs. After)

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 **received** Balsakhi

with
test
scores
of...



Children who **did not** receive Balsakhi

2 - Simple difference

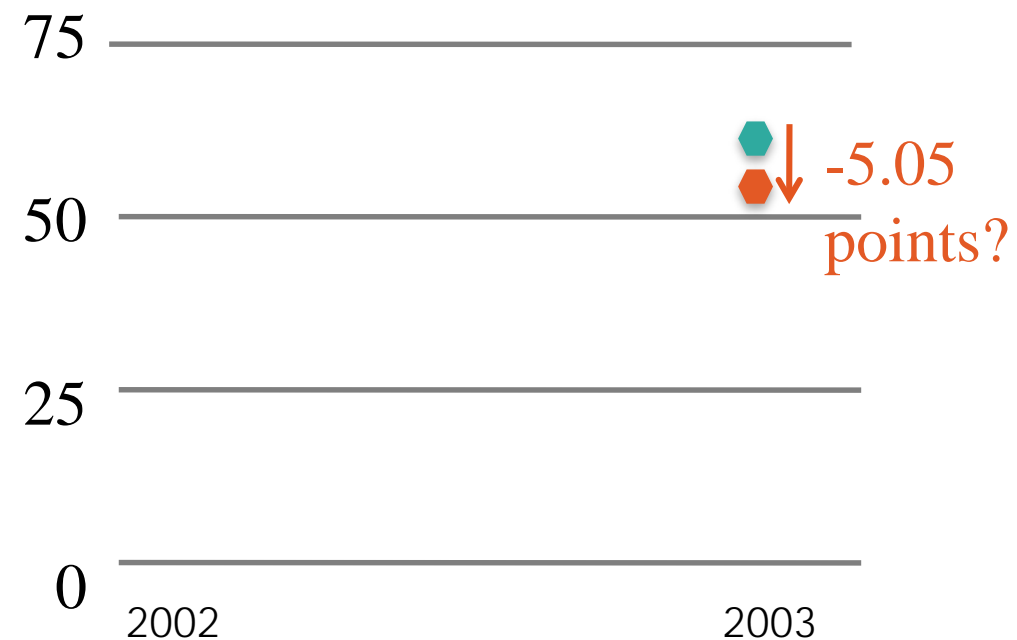
Average score for children with a Balsakhi	51.22
Average score for children without a Balsakhi	56.27
Difference	-5.05

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

What would have happened without a Balsakhi?

Method 2: Simple Comparison

Impact = -5.05 points?



3 – Difference-in-Differences

Compare gains in test scores of...



Children who **received** Balsakhi

with gains in test scores of...



Children who **did not receive** 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-in-Differences			6.82

Question: Under what conditions can this difference (6.82) be interpreted as the impact of the Balsakhi program?

4 – Other Methods

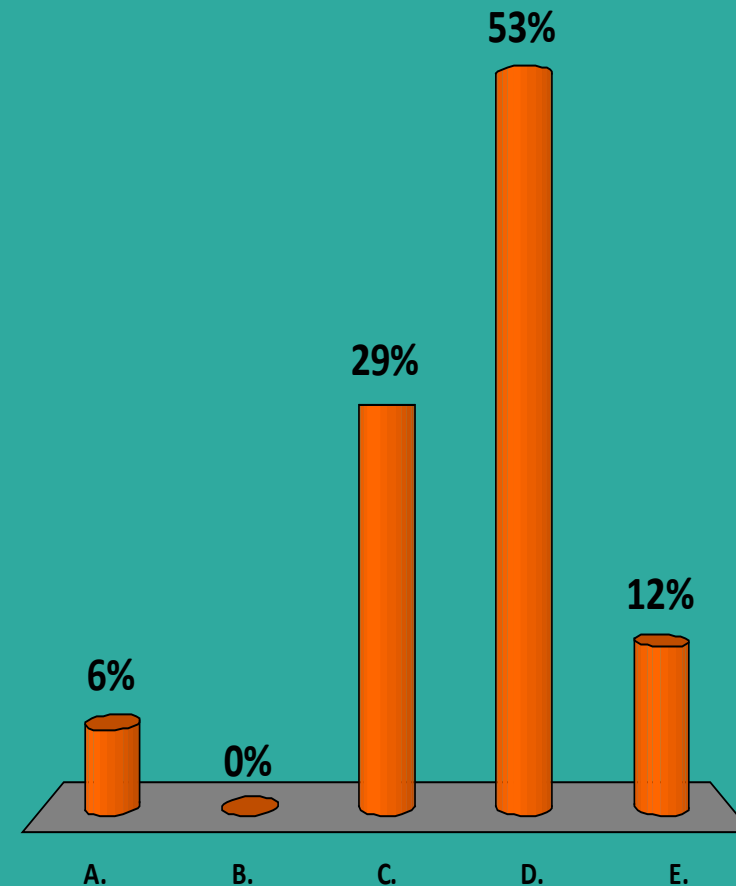
- There are more sophisticated non-experimental methods to estimate program impacts:
 - Regression
 - Matching
 - Instrumental Variables
 - Regression Discontinuity
- These methods rely on being able to “mimic” the counterfactual **under certain assumptions**
- **Problem:** Assumptions are not testable

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?

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

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



Impact Estimates of Balsakhi - Summary

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

*: Statistically significant at the 5% level

Bottom Line: Which method we use matters

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

4. Conclusions

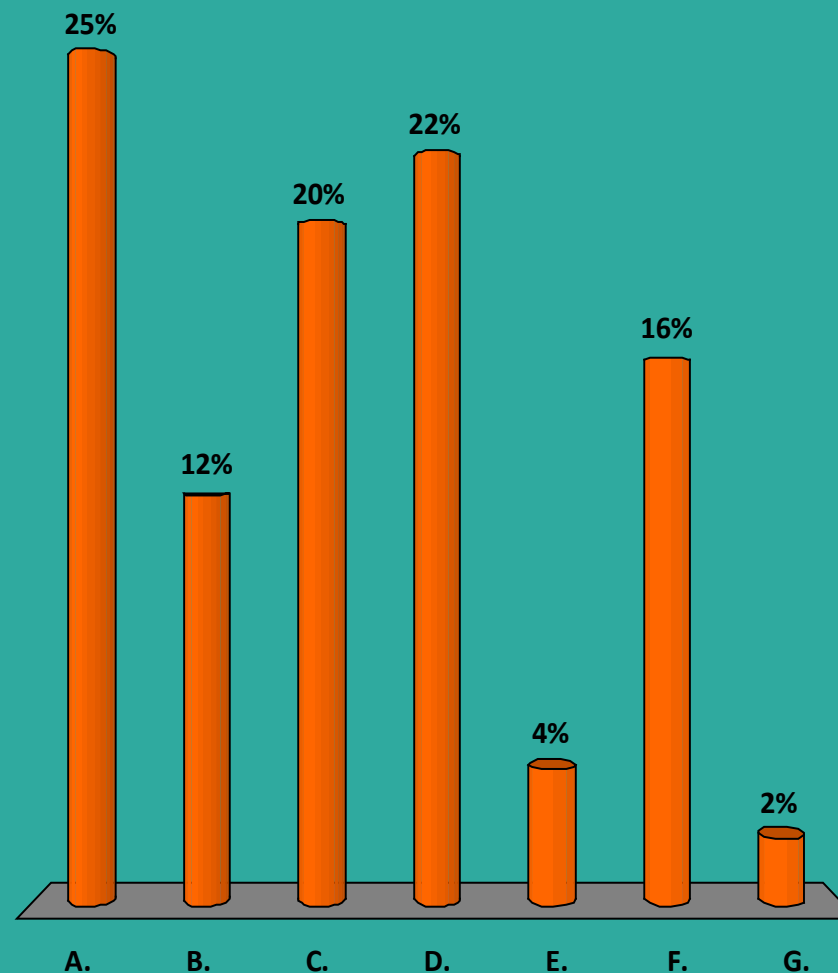


Conclusions - Why Randomize?

- 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

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, but not why or how it occurred (i.e. it is a black box)



THANK YOU!

