ABDUL LATIF JAMEEL
Poverty Action Lab

TRANSLATING RESEARCH INTO ACTION

Impact Evaluation: Why randomize?

Dan Levy

Harvard Kennedy School

Abdul Latif Jameel Poverty Action Lab

Course Overview

1. Why evaluate? What is evaluation?
2. Outcomes, indicators and measuring impact
3. **Impact evaluation – why randomize?**
4. How to randomize
5. Sampling and sample size
6. Implementing an evaluation
7. Analysis and inference

Lecture Overview

I. Background

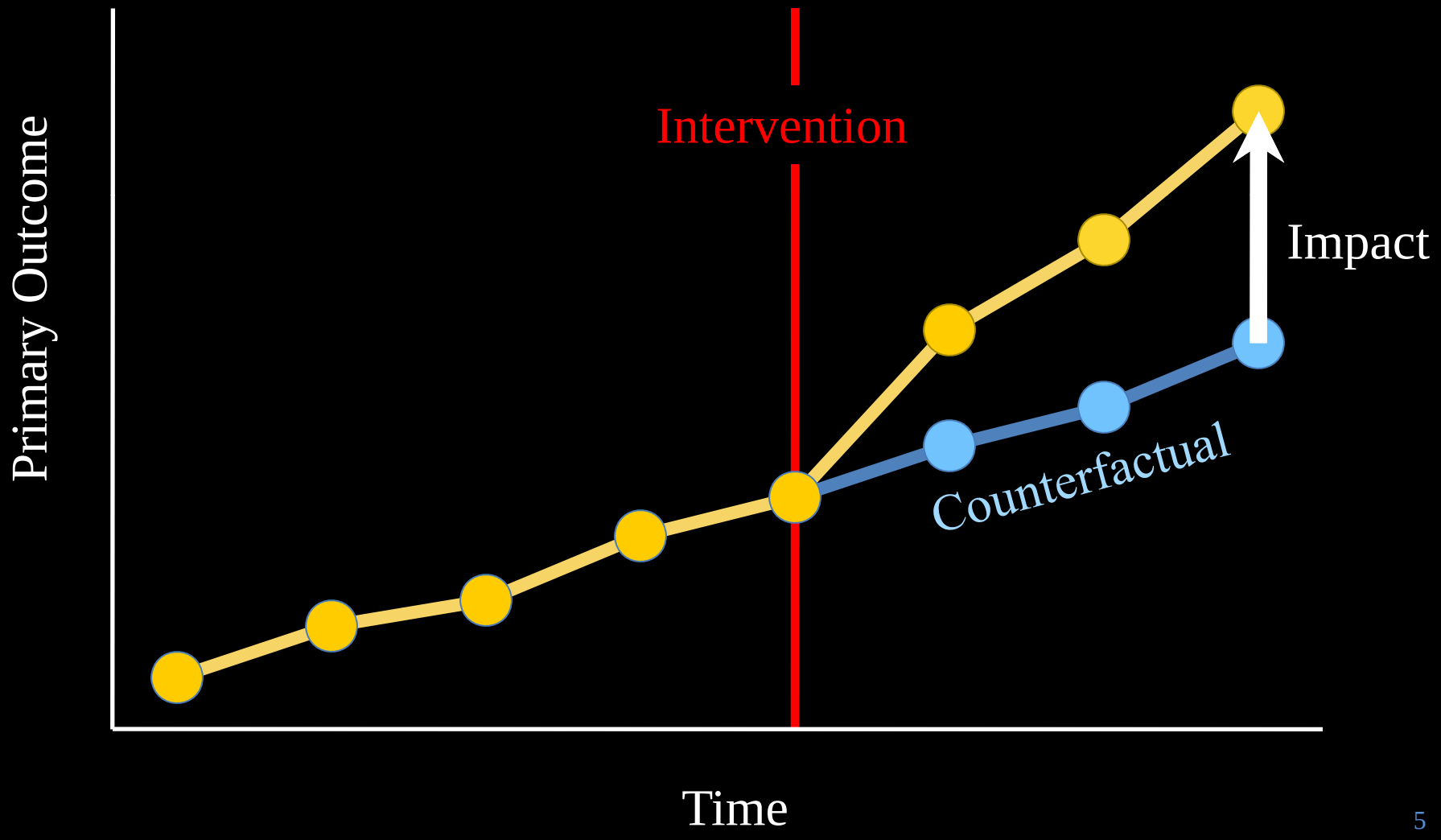
II. What is a randomized experiment?

III. Why randomize?

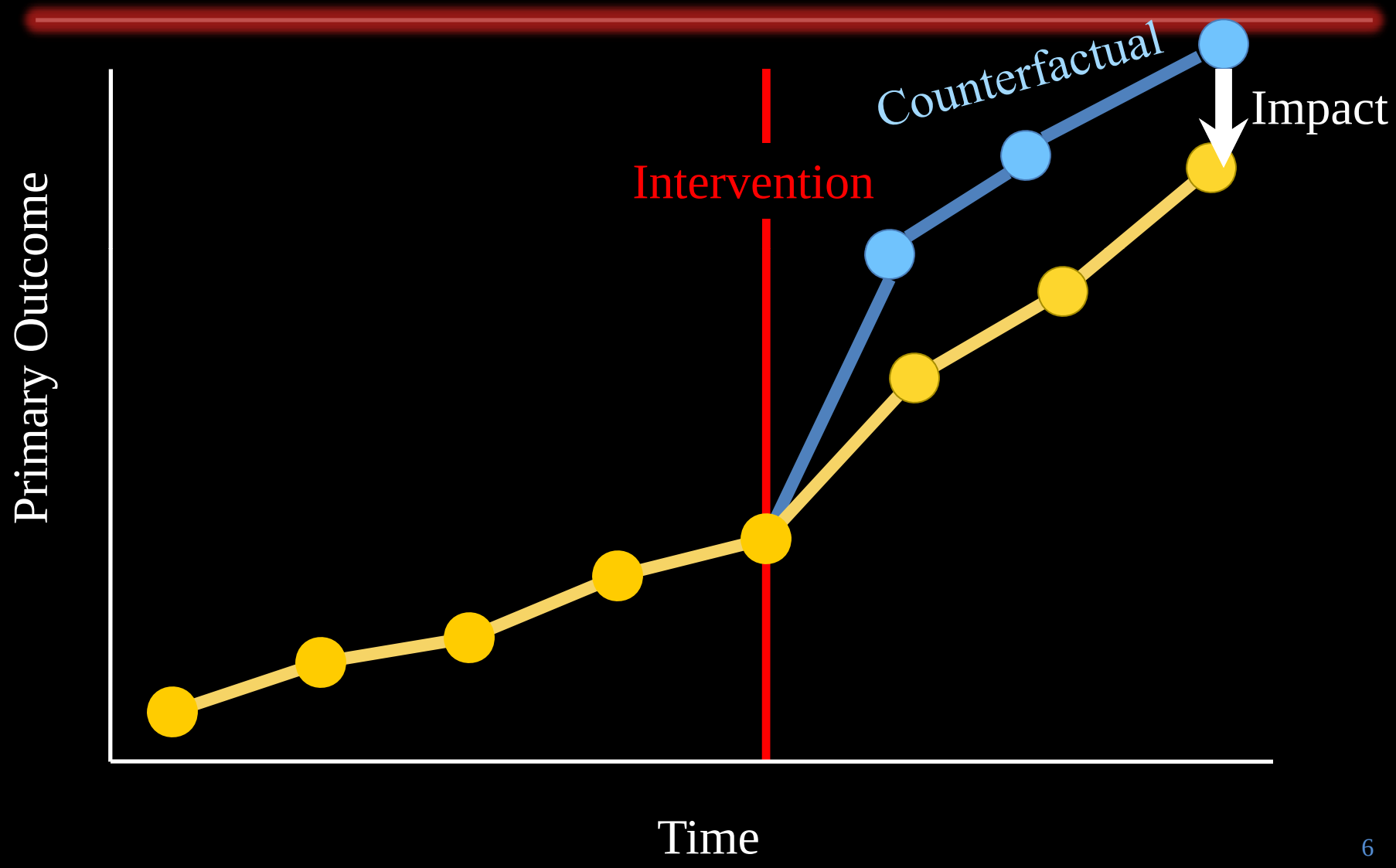
IV. Conclusions

I - Background

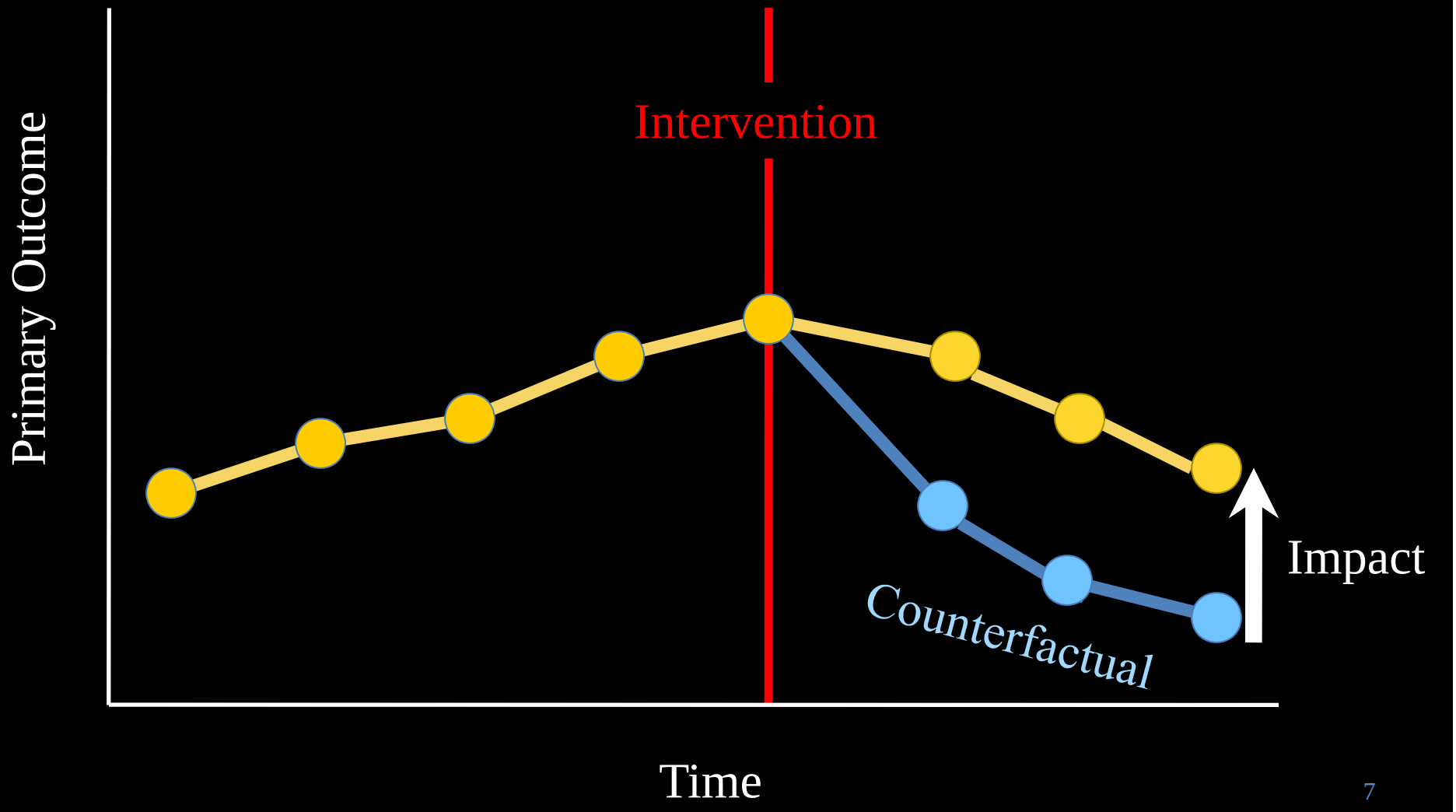
Impact: What is it?



Impact: What is it?



Impact: What is it?



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"***)

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

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?

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 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?

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

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?

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)

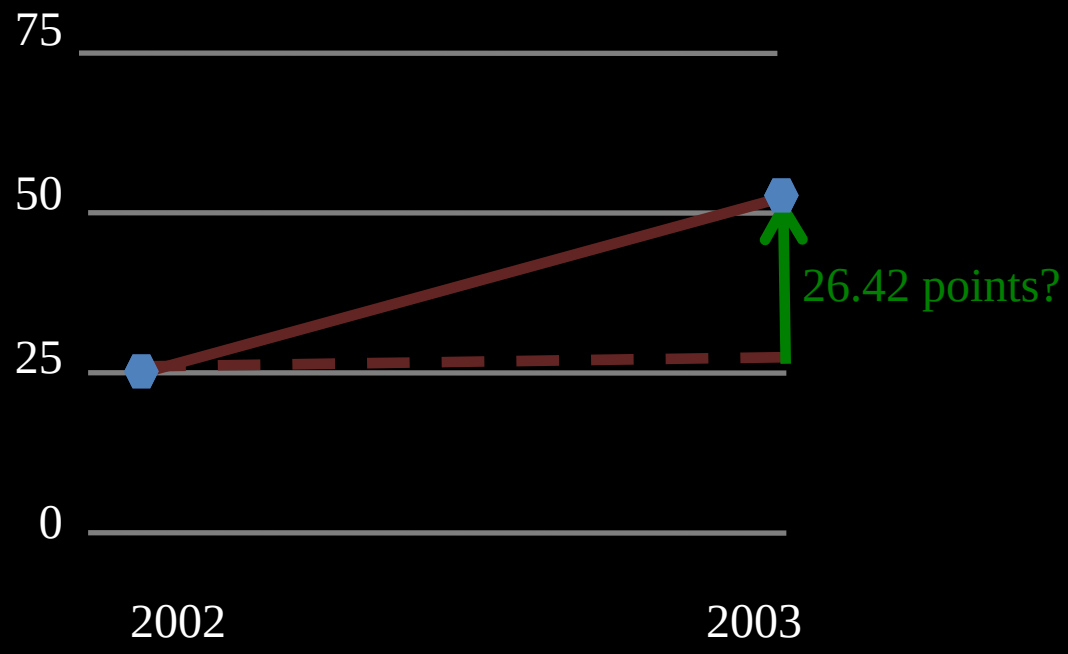
Average **post-test** score for children with a balsakhi	51.22
Average **pretest** score for children with a balsakhi	24.80
Difference	**26.42**

- 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

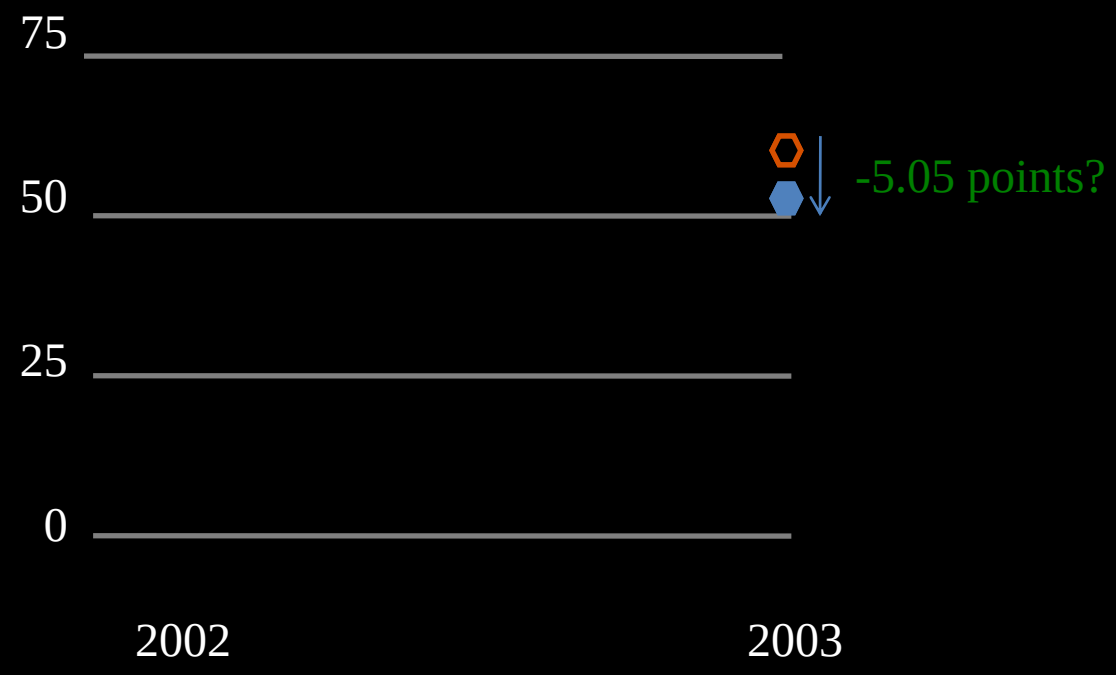
Average score for children with a balsakhi	**51.22**
Average score for children without a balsakhi	56.27
Difference	-5.05

- 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
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?

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?

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!

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	0.88*

*: Statistically significant at the 5% level

Example #3: A voting campaign in the USA

Courtesy of Flickr user theocean

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

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

Questions/Comments?

Dan Levy
Harvard Kennedy School

Resource: Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs

Dr. Rachel Glennerster, Prof. Abhijit Banerjee, Prof. Esther Duflo