Quantitative Impact Evaluation Methods

Ali Asjad Naqvi
Research and Training Director
Center for Economic Research in Pakistan (CERP)

Teaching Evaluation in South Asia (TESA)
July 25, 2013
Quantitative Evaluation Methods

- Pre-Post
- Simple Difference
- Difference-in-Difference
- Multivariate Regression
- Statistical Matching
- Instrumental Variables
- Regression Discontinuity
- Randomized Evaluations
Quantitative Evaluation Methods

- Pre-Post
- Simple Difference
- Difference-in-Difference
- Multivariate Regression
- Statistical Matching
- Instrumental Variables
- Regression Discontinuity
- Randomized Evaluations
Example: Pratham’s Balsakhi Program
Baseline findings

- Many children in 3rd and 4th standard were not even at the 1st standard level of competency
- 44% of kids between 7-12 could not read a basic paragraph
- 50% could not do simple subtraction despite being enrolled in school
- Only 19.5% students in Grade 3 could correctly answer Grade 1 math problems

What were the problems?
What were the problems?

- Large class sizes?
- Low competency levels in higher classes?
- Social gaps?
  - Amongst students
  - Between students and teachers
Proposed solutions

- Hire local teachers \((Balsakhis)\)
  - Given them training in Hindi, Maths, English

- Identify lowest performing students from grade 3 and 4:
  - Take them out of classroom for two hours
  - Ask \(Balsakhis\) to teach them
Proposed implementation

Study design:

- We want to look at the impact of providing Balsakhis on learning outcomes.
- Implement the program in all public schools in the state of Gujrat in India.

Let’s evaluate this using different methods.
Method 1: Pre vs Post

- Take the students **enrolled** in the *Balsakhi* program
- Look at their scores at the **start** and **end** of the Balsakhi program
METHOD 1: PRE VS POST

Average test scores of Balsakhi students

<table>
<thead>
<tr>
<th>Start of program</th>
<th>End of program</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.8</td>
<td>51.22</td>
</tr>
</tbody>
</table>

26.42

QUESTION: Under what conditions can the difference of 26.42 be interpreted as the impact of the Balsakhi program?
Method 1: Pre vs Post

Average test scores of Balsakhi students

Start of program | End of program

Impact = 26.42 points?

What would have happened without the Balsakhi program?
**What is 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 or the **counterfactual**
Impact: What is it?

A. Positive
B. Negative
C. No impact
D. Don’t Know

Primary Outcome

Intervention

40%  40%

0%  20%  20%
**What is Impact?**

![Diagram showing the concept of Impact with timelines and primary outcomes.](image)
**Impact: What is it?**

- A. Positive
- B. Negative
- C. No impact
- D. Don’t Know

![Chart showing primary outcome vs. time with Intervention and Counterfactual trends, and bar chart displaying percentages of positive, negative, no impact, and don’t know responses.](chart.png)

- Positive: 9%
- Negative: 64%
- No impact: 9%
- Don’t know: 18%
Method 2: Simple Difference

- Divide the population into **two** groups:
  - One group enrolled in Balsakhi program (Treatment)
  - One group not enrolled in Balsakhi program (Control)

- Compare test score of these two groups at the **end** of the program.
**METHOD 2: SIMPLE DIFFERENCE**

**Average test scores end of program**

<table>
<thead>
<tr>
<th></th>
<th>Not enrolled in program</th>
<th>Enrolled in program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>56.27</td>
<td>51.22</td>
</tr>
<tr>
<td>Difference</td>
<td>-5.05</td>
<td></td>
</tr>
</tbody>
</table>

**QUESTION:** Under what conditions can the difference of **-5.05** be interpreted as the impact of the Balsakhi program?
Method 3: Difference-in-difference

- Divide the population into two groups:
  - One group enrolled in Balsakhi program (Treatment)
  - One group not enrolled in Balsakhi program (Control)

- Compare test score of these two groups at the start and at the end of the program.
Method 3: Difference-in-difference

Average test scores

<table>
<thead>
<tr>
<th></th>
<th>Start of program</th>
<th>End of program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled in Balsakhi program</td>
<td>24.8</td>
<td>51.22</td>
</tr>
<tr>
<td>Not enrolled in Balsahki program</td>
<td>36.67</td>
<td>56.27</td>
</tr>
</tbody>
</table>
Method 3: Difference-in-difference

Average test scores

Start of program
End of program

- Enrolled in Balsakhi program
- Not enrolled in Balsahki program

26.42
19.60
**METHOD 3: DIFFERENCE-IN-DIFFERENCE**

**QUESTION:** Under what conditions can the difference of 6.82 be interpreted as the impact of the Balsakhi program?
Method 4: Regression analysis

- Divide the population into two groups:
  - One group enrolled in Balsakhi program
  - One group not enrolled in Balsakhi program

- Compare test score of these two groups at the start and at the end of the program.

- **Control** for additional variables like gender, class-size

- Post-test = $\beta_0 + \beta_1 \text{Pre-test} + \beta_2 \text{Gender} + \beta_3 \text{Class-size} + \beta_4 \text{Balsakhi} + e$
QUESTION: Under what conditions can the coefficient of 1.92 be interpreted as the impact of the Balsakhi program?
Impact of Balsakhi Program

(1) Pre-post *  (2) Simple Difference *  (3) Difference-in-Difference *  (4) Regression with controls

* Significant at 5% level
## Summary of Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Comparison</th>
<th>Works only if…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Post</td>
<td>Program participants before program</td>
<td>Nothing else was affecting outcome</td>
</tr>
<tr>
<td>Simple Difference</td>
<td>Individuals who did not participate (data collected after program)</td>
<td>Non-participants are exactly equal to participants</td>
</tr>
<tr>
<td>Differences-in-Difference</td>
<td>Same as above + data collected before and after</td>
<td>If two groups have exactly the same trajectory over time</td>
</tr>
<tr>
<td>Regression</td>
<td>Same as above + additional “explanatory” variables</td>
<td>Omitted variables do not affect results</td>
</tr>
</tbody>
</table>
5. Other Methods

- There are more sophisticated non-experimental and quasi-experimental methods to estimate program impacts:
  - Statistical 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

Groups are statistically identical before the programs, any difference can be attributed to the program.

Individuals, clients, firms, villages are randomly selected to receive the treatment, while other units serve as a comparison.
Basic set-up of a randomized evaluation

Total Population

Target Population

Not in evaluation

Evaluation Sample

Random Assignment

Treatment Group

Control Group
Randomly sample from area of interest
Randomly **sample** from area of interest

Randomly **assign** to treatment and control

Randomly **sample** from both treatment and control
Randomization Design

- Population = all schools in case villages
- Target population: weakest students in all of these schools
- Stratify on three criteria:
  - Pre-test scores
  - Gender
  - Language
- Give 50% of them the Balsakhi program
Context and Partner

- Abhijeet Banerjee, Esther Duflo, Shawn Cole, Leigh Linden

- 122 Primary municipal schools in Vadodara and 77 schools in Mumbai (Western India)

- 2002 & 2003: Two academic years

- ~17,000 children
Impact of Balsakhi program

(1) Pre-post *  (2) Simple Difference *  (3) Difference-in-Difference *  (4) Regression with controls

(5) RCT = 5.87

* Significant at 5% level
Conclusions

- The method you chose will determine results
  - Know the limitations of each method

- Understand the importance of the counterfactual

- RCTs are considered the most rigorous in evaluations since they always ensure a counterfactual
  - Pure impact of interventions
## Summary of Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Comparison</th>
<th>Works only if…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Post</td>
<td>Program participants before program</td>
<td>Nothing else was affecting outcome</td>
</tr>
<tr>
<td>Simple Difference</td>
<td>Individuals who did not participate (data collected after program)</td>
<td>Non-participants are exactly equal to participants</td>
</tr>
<tr>
<td>Differences-in-Difference</td>
<td>Same as above + data collected before and after</td>
<td>If two groups have exactly the same trajectory over time</td>
</tr>
<tr>
<td>Regression</td>
<td>Same as above + additional “explanatory” variables</td>
<td>Omitted variables do not affect results</td>
</tr>
<tr>
<td>Randomized Evaluation</td>
<td>Participants randomly assigned to control group</td>
<td>The two groups are statistically identical on observed and unobserved characteristics</td>
</tr>
</tbody>
</table>