EMPath
Economic Mobility Pathways

Presents

Disrupting The Poverty Cycle Conference 2016

Breakthrough Interventions and Outcomes

October 6-7, 2016
University of Massachusetts - Boston
Evaluating Social Programs

Jason Bauman and Julia Chabrier
J-PAL North America (MIT)
J-PAL: 750+ ongoing and completed projects in 65+ countries
J-PAL’s network of affiliated researchers
J-PAL’s mission is to reduce poverty by ensuring that policy is informed by scientific evidence.

**EVALUATIONS**

J-PAL researchers conduct randomized evaluations to test and improve the effectiveness of programs and policies aimed at reducing poverty.

**CAPACITY BUILDING**

Through training courses, evidence workshops, and research projects, J-PAL equips policymakers and practitioners with the expertise to carry out their own rigorous evaluations.

**POLICY OUTREACH**

J-PAL affiliates and staff analyze and disseminate research results and build partnerships with policymakers to ensure policy is driven by evidence and effective programs are scaled up.
I. Why Evaluate?

II. What is Evaluation?

III. Measuring Impact

IV. Group Work: Designing an Impact Evaluation

V. Case Study: Health Care Hotspotting
I. Why Evaluate?
II. What is Evaluation?
III. Measuring Impact
IV. Group Work: Designing an Impact Evaluation
V. Case Study: Health Care Hotspotting
What can prevent violence?

• 610 Chicago Public School students were shot between September 2008 and April 2010

• “Nothing stops a bullet like a job”
  – Fr. Greg Boyle, Homeboy Industries

• Limited causal evidence that employment programs reduce crime
One approach: Chicago’s youth programs

- $36 million budget for after-school programs and summer jobs
- A summer job might reduce violence
  - Soft skills
  - “Keeping busy”
  - Income
One Summer Chicago Plus (OSP)

Experimental program for 8th through 12th graders in 2012
- Government and non-profit minimum-wage jobs
- 25 hours/week
- Adult job mentor
- 1-day training
OSP reduced violent crime

Violent Crime Arrests

Average # arrests per youth

Control Mean: 0.091
Treatment Mean: 0.051

43% decrease
The policy impact of this finding

• $10 million from private philanthropists to expand the program:
  – 2,000 youth in 2015
  – 3,000 youth in 2016
  – 4,000 youth in 2017
I. Why Evaluate?

II. What is Evaluation?

III. Measuring Impact

IV. Group Work: Designing an Impact Evaluation

V. Case study: Health Care Hotspotting
What is evaluation?
Program evaluation

Evaluation

Program Evaluation

Impact Evaluation

J-PAL | Evaluating Social Programs
Programs and their evaluations: Where do we start?

Program Creation
- Start with a problem
- Verify that the problem actually exists
- Generate a theory of why the problem exists
- Design the program
- Think about whether the solution is cost effective

Program Evaluation
- Start with a question
- Verify the question hasn’t been answered
- State a hypothesis
- Design the evaluation
- Determine whether the value of the answer is worth the cost of the evaluation
Components of program evaluation

<table>
<thead>
<tr>
<th>Component</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needs Assessment</td>
<td>• What is the problem?</td>
</tr>
<tr>
<td>Theory of Change</td>
<td>• How, in theory, does the program fix the problem?</td>
</tr>
<tr>
<td>Process Evaluation</td>
<td>• Does the program work as planned?</td>
</tr>
<tr>
<td>Impact Evaluation</td>
<td>• Were its goals achieved?</td>
</tr>
<tr>
<td></td>
<td>The magnitude?</td>
</tr>
<tr>
<td>Cost Effectiveness</td>
<td>• Given magnitude and cost, how does it compare to alternatives?</td>
</tr>
</tbody>
</table>
Theory of change

• Theory of change is a structured way of thinking—a road map of where we are going (results) and how we are getting there (process)
Implementation failure vs. theory failure

Successful intervention

Input → Activities → Outputs → Outcomes → Goal

Implementation failure

Input → Activities → Outputs → Outcome → Goal

Theory failure

Input → Activities → Outputs → Outcome → Goal
I. Why Evaluate?

II. What is Evaluation?

III. Measuring Impact

IV. Group Work: Designing an Impact Evaluation

V. Case Study: Health Care Hotspotting
How impact differs from process

• **Process question**: What happened?

• **Impact question**: What happened relative to what would have happened without the program?

• **Counterfactual**: What would have happened without the program
Impact: What is it?

Program starts

Time

Primary Outcome

Impact

Counterfactual
Impact: What is it?

Impact: What is it?

Program starts

Counterfactual

Impact

Primary Outcome

Time

J-PAL | Evaluating Social Programs
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)

**Problem**: Counterfactual cannot be observed

**Solution**: We need to “mimic” or construct the counterfactual
Selecting the comparison group

• **Goal:** Attribute differences in outcomes between the group of participants and the comparison group to the program (and not to other factors)

• **Strategy:** Select a group that most closely approximates the group of participants in all ways except one—their exposure to the program being evaluated
Evaluation methods

• What are some methods you’ve used or considered using to evaluate the impact of your work?
Methods as tools

- Pre-post
- Simple Difference
- Difference-in-Difference
- Regressions
- Randomized Evaluation
Evaluating a tutoring program
Tutoring program: Background

• Implemented by Pratham, an NGO in India
• Program provided community tutors to help at-risk children with school work
• In Vadodara, an Indian city, the tutoring program was run in government primary schools in 2002-2003
• Teachers decided which children would get the tutor
1. Pre-Post

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average pre-test score for children with a tutor</td>
<td>24.80</td>
</tr>
<tr>
<td>Average post-test score for children with a tutor</td>
<td>51.22</td>
</tr>
<tr>
<td>Difference</td>
<td>26.42</td>
</tr>
</tbody>
</table>
What would have happened without tutors?

Method 1: Pre-Post
Impact = 26.42 points?
2. Simple Difference

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score for children</td>
<td>51.22</td>
</tr>
<tr>
<td>with a tutor</td>
<td></td>
</tr>
<tr>
<td>Average score for children</td>
<td>56.27</td>
</tr>
<tr>
<td>without a tutor</td>
<td></td>
</tr>
<tr>
<td>Difference</td>
<td>-5.05</td>
</tr>
</tbody>
</table>
What would have happened without tutors?

Method 2: Simple Difference
Impact = -5.05 points?
3. Difference-in-Differences

Compare gains in test scores of...

Children who got tutors with...

Children who did not need tutors

With test scores of...
## 3. Difference-in-Differences

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Post-test</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score for children with a tutor</td>
<td>24.80</td>
<td>51.22</td>
<td>26.42</td>
</tr>
<tr>
<td>Average score for children without a tutor</td>
<td>36.67</td>
<td>56.27</td>
<td>19.60</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td>6.82</td>
</tr>
</tbody>
</table>
What would have happened without tutors?

Method 3: Difference-in-Differences
Impact = 6.82 points?
4. Randomized Evaluation

- Identify all classrooms that are eligible to have a tutor
- Randomly assign some classrooms to be a part of the tutoring programs
- Measure difference between classes randomly assigned to have access to a tutor and those randomly assigned to not have access
Impact of tutoring: Summary

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-Post</td>
<td>26.42*</td>
</tr>
<tr>
<td>(2) Simple Difference</td>
<td>-5.05*</td>
</tr>
<tr>
<td>(3) Difference-in-Differences</td>
<td>6.82*</td>
</tr>
<tr>
<td>(4) Randomized Evaluation</td>
<td>5.87*</td>
</tr>
</tbody>
</table>

* Statistically significant at the 5% level
Why randomize?

- Mathematically, it can be shown that random assignment makes it very likely that we are making an apples-to-apples comparison.
How to randomize

1. Identify eligible participants
2. Random lottery
3. Treatment
4. Control
5. Intervention
6. Measure outcomes
When to consider randomization

• When your program is...
  – Over-subscribed
  – Being rolled out ("Smart-piloting")
  – Expanding (e.g., moving into a new location or service area)
  – Adding a new component
I. Why Evaluate?

II. What is Evaluation?

III. Measuring Impact

IV. Group Work: Designing an Impact Evaluation

V. Case Study: Health Care Hotspotting
Designing an impact evaluation

- Imagine that your organization is rolling out a new program that offers in-person mentoring to high school seniors at risk of not applying to college

- How might you design an impact evaluation using one of these methods:
  - Pre-post
  - Simple difference
  - Randomized evaluation
Designing an Impact Evaluation

1. What is the question that you would like to answer with an impact evaluation?
2. Who is the target population for the program?
3. How would participants be selected to receive the program?
4. Who would serve as the comparison group for estimating the program’s impact?
5. What are some of the advantages of disadvantages of this evaluation design?
To Learn More…


I. Why Evaluate?
II. What is Evaluation?
III. Measuring Impact
IV. Group Work: Designing an Impact Evaluation
V. Case Study: Health Care Hotspotting
Health care hotspotting

- Camden Coalition of Health Care Providers’ Link2Care program serves “super-utilizers” of health care system
- Care teams perform home-visits, accompany patients to initial doctor visits, and help patients enroll in social-service programs
Identifying eligible participants

• Capacity constraint makes randomization feasible
• Targeted sample size: 400 controls and 400 intervention
Recruitment, consent, and randomization

- Health Information Exchange provides daily report of patients with 2+ hospital admissions in the last 6 months
- Camden Coalition staff introduce the program, obtain consent, and randomize using survey software on tablets
Service Delivery and Process Monitoring

- Real-time data records services delivered to treatment group, including when their first home visit occurred and whether they saw their primary care physician.
Outcomes

• Real-time analysis of primary outcome (hospital readmissions) using Health Information Exchange

• Measure longer-term outcomes in administrative data
  • Health care use: Medicare and Medicaid claims
  • Participation in social programs: NJ Department of Health and Human Services TANF and SNAP data, Camden Housing Authority data
  • Criminal justice involvement: NJ Department of Corrections data
  • Earnings and employment: NJ and PA Department of Labor unemployment insurance records
  • Mortality: Social Security Administration and NJ Vital Statistics
How can we be a resource?

Research resources on www.povertyactionlab.org

• Guide to Using Administrative Data
• Introduction to Evaluations
• The Goldilocks Toolkit (by Innovations for Poverty Action)

If your organization is interested in a randomized evaluation, please feel free to reach out!
Evaluating Social Programs

- Jason Bauman
  - jbauman@mit.edu

- Julia Chabrier
  - chabrier@mit.edu

- J-PAL NA
  - www.povertyactionlab.org/na
  - @JPAL_NA