Joshua Blumenstock, UC Berkeley

Social Impact Evaluation

Outline: Impact Evaluation

- What is it?
- Why do it?
- How to do it?
  - The easy way
  - The hard way
- Doing it: more resources

What is impact evaluation?

- Measuring the effect of a "treatment" on a beneficiary population
  - Treatment: project, intervention, policy, etc.
- Accounting/Controlling for all other factors that might have affected the target population during the program period
- "IMPACT" = portion of an outcome that can be attributed directly to the treatment
- Difference between what happened with the program, and what would have happened without, to the same target group

Impact Evaluation vs. monitoring?

- Impact evaluation to measure effectiveness (focusing on final outcomes)
- Monitoring & Evaluation to track implementation efficiency (focusing on use of inputs)
- Impact Evaluation:
  - What was the effect of the program on outcomes?
  - How would outcomes change under alternative program designs?
  - Does the program impact people differently (e.g. females, poor, minorities)?
  - Is the program cost-effective?
- Process Evaluation / Monitoring:
  - Is program being implemented accurately and efficiently?
  - Is program targeting the right population?
  - Are outcomes moving in the right direction?

Why do impact evaluation?

- Interventions and policies seem obvious…
  - Kenyan students don't have textbooks!
  - Glewwe et al (2009) provided free textbooks
  - Only helps top students, English speakers
- What about other interventions?
  - There aren't enough teachers!
  - There isn't a way to get to school!
  - Kids are healthy at home!
  - No food at school!

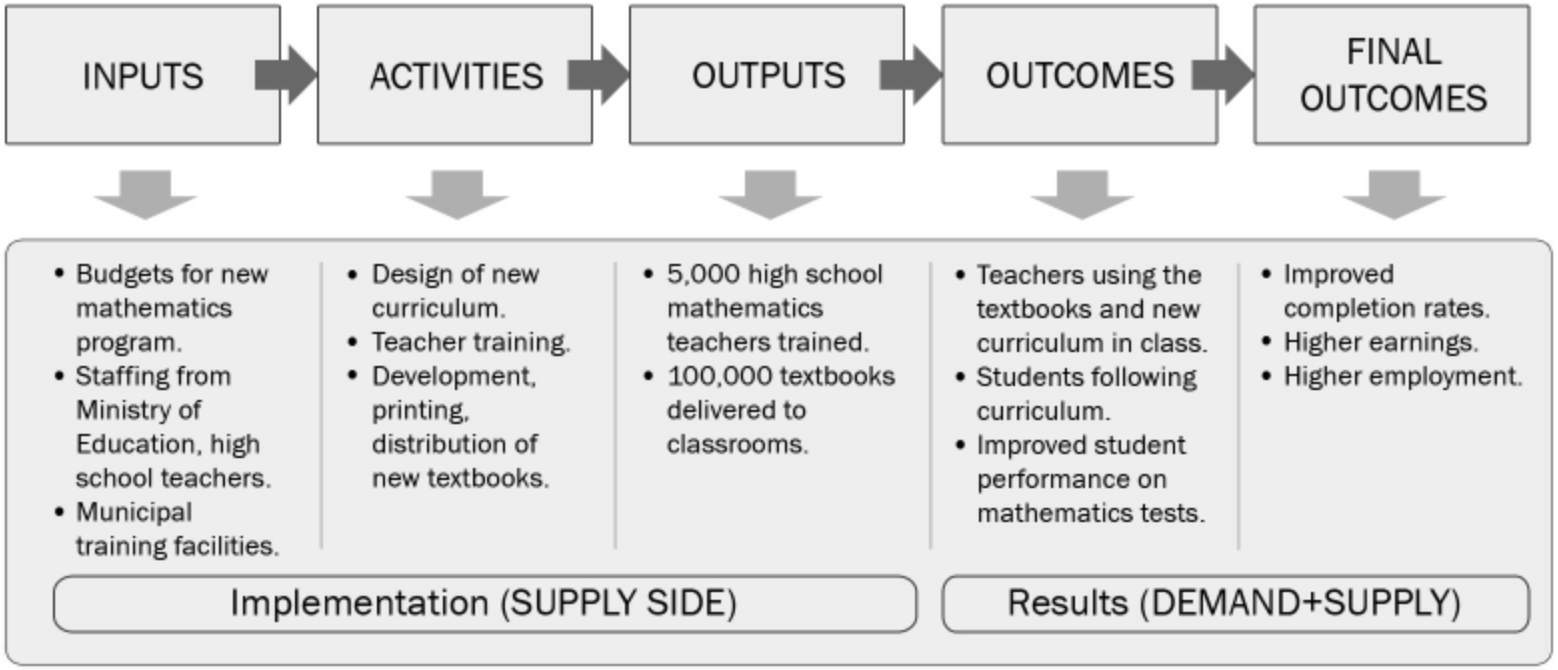
Why do impact evaluation?

- If you need evidence that a program works
  - Does the program actually improve outcomes? i.e., does it improve welfare for the people most in need?
    - Accountability to civil society
    - Accountability to donors
    - Ability to target a limited budget
- If you want to improve the program over time
  - Results-based management
  - Cut ineffective components or programs
  - Cost-effectiveness
- If you want to scale up but need proof of concept
- If you want to compare it to other programs

Outline: Impact Evaluation

- What is it?
- Why do it?
- How to do it?
- No really, how to do it?
- Doing it: more resources

The Results Chain



Sometimes the hardest part…

- … Is Defining your outcome (Gertler ch.2)
  - Specific: measure required info closely
  - Measurable: can actually be obtained
  - Attributable: linked to project
  - Realistic: timely, reasonable cost
  - Targeted: to population of interest
- Find indicators throughout results chain

Fundamental challenge of impact evaluation

- We want to be able to say if this program affected an outcome (increased earnings, employment, happiness)
  - The Causal Effect
- Need to know what earnings, employment, happiness would have been in the absence of the program.
  - The "Counterfactual"
  - How is this different from a "control group?"
- But we can never directly observe the counterfactual!
- So why not just compare treated with untreated?

Correlation vs. Causation

- A SMS-based service to remind mothers to wash their hands. This is the intervention.
- Does this service increases infant health outcomes?
  - Compare people with service to those without
    - Simple averages: $Y = \alpha + \beta T + \varepsilon$
    - multiple regression: $Y = \alpha + \beta T + \delta Z + \varepsilon$
    - matching
  - Compare same people before/after service

Simple comparisons: the problem

- Compare people with service to those without
  - E.g. people in capital vs. people in rural areas, people in capital who enrolled with people in capital who didn't
  - Control for Z factors. Problem: Why did the enrollees enroll?
    - Z can be easy to measure: Location, age, demographics
    - Z can be hard to measure
      - motherly instincts, motivation, resourcefulness, cleanliness
  - Selection Bias: Apples vs. oranges
- Compare people before service to same people after
  - E.g. first child (before service) vs. second child (after)
  - Is difference in outcome due to the project?
  - Problem: Many other things probably changed over time
  - Omitted Variable Bias: young apples vs. old apples?

Three big problems

- Omitted Variable Bias: If you're leaving out something from your analysis…
  - …. and that something is correlated with the treatment
- Selection Bias: Special case of OVB
  - The people / groups / places being treated differ systematically from those who aren't being treated.
  - They are "selecting into treatment".
  - So the untreated aren't a valid counterfactual for the treated group.
- Reverse causality: "Special case of Selection"

Constructing a counterfactual

- The easy way: randomization
- The hard way
  - Regression Discontinuity
  - Difference-in-Difference
  - Matching
  - Instrumental variables

How to do it: the easy way

- Randomization
  - Gold-standard in impact evaluation
  - Use a lottery to give all people an equal chance of being in control or treatment groups
    - Doesn't have to be 50-50
  - With a large enough sample, it guarantees that all factors/characteristics will be equal between groups, on average
  - Only difference (on average) between 2 groups is the treatment
    - Omitted Variables, Selection bias no longer a problem
  - It is a fair and transparent method to allocate scarce resources

How to randomize, pt.1

- Define population of interest
  - Cost vs. Statistical Power (see Duflo et al. sec. 4)
- Design the experiment carefully
- Randomly assign people to treatment or control
- Collect baseline data
- Verify that assignment looks random

How to randomize, pt.2

- Monitor process so that integrity of experiment is not compromised
- Collect follow-up data for treatment and control groups in identical ways
- Estimate program impacts by comparing mean outcomes of treatment group vs. mean outcomes of control group
- Assess whether program impacts are statistically significant

Validity

- Internal validity
  - Relates to ability to draw causal inference, i.e. can we attribute our impact estimates to the program and not to something else
  - Random Assignment
- External validity
  - Relates to ability to generalize to other settings of interest, i.e. can we generalize our impact estimates from this program to other populations, time periods, countries, etc?
  - Random Sampling

Randomization: some issues

- Identify target population, then randomize treatment and control within it
  - Individual Level
  - Community / Village Level
- Things to consider
  - What's population to make inferences about?
  - Spillovers / Externalities (Kremer & Miguel 2002)
    - Positive examples?
    - Negative examples?
    - How do these affect estimates of "treatment" effect?
  - Costs & Statistical Power (e.g. transportation costs, staff costs)

Randomization: Practical approaches

- Permanent treatment / control design
- Multiple treatment arms
- Phased roll out
- Randomized promotion or encouragement design

Randomization: remember…

- Impact evaluation needs data!
- Baseline Data (before randomization)
- Follow-up Data (after randomization)
- What data will you collect?
- How accurate will it be? (e.g. income data hard to measure)
- HAVE to gather data on control observations

Ethical issues

- Can you ethically withhold treatment from the control group?
- What steps will you take to monitor the effect of treatment?
- Who will review your proposal to examine ethical concerns?

How to do it: the hard way

- What if it's impossible to randomize
  - Too late
- Find a counterfactual
- Find a "natural experiment"
  - Where you can say, after accounting for X, the treatment assignment was "as good as random"

Regression Discontinuity

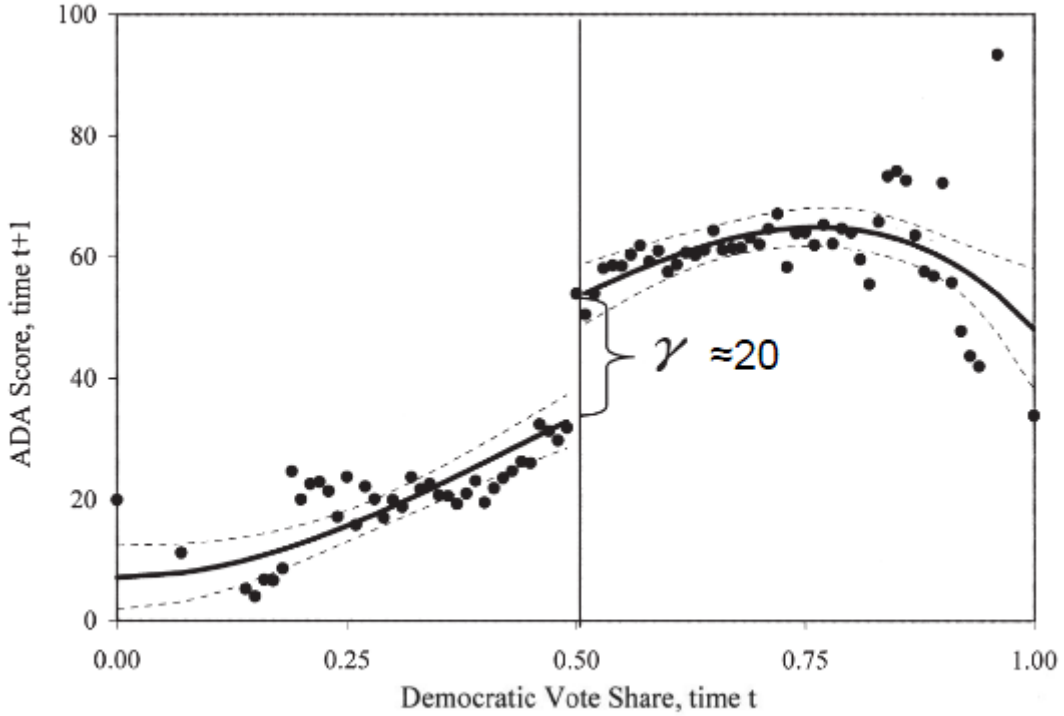


FIGURE I
Total Effect of Initial Win on Future ADA Scores: γ

Difference-in-Difference

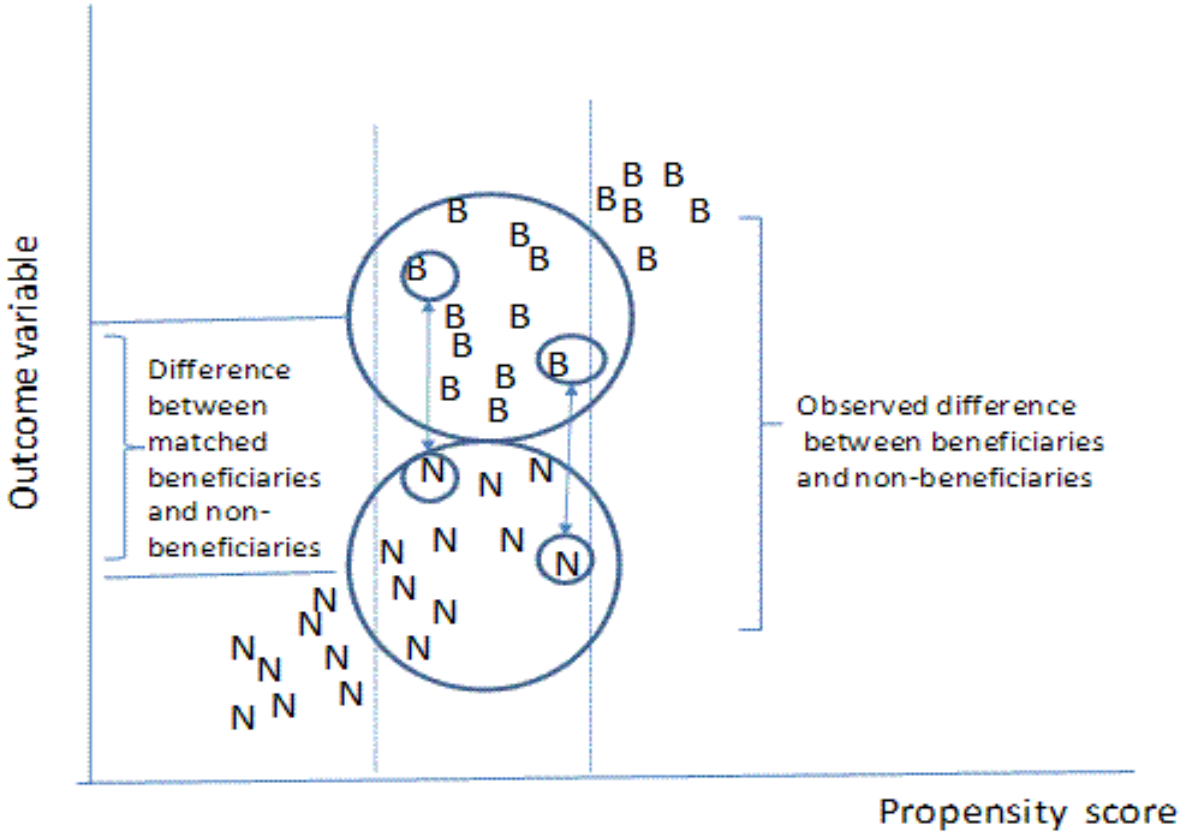
TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

Matching

Figure 1. A graphical representation of matching on the propensity score



Further resources

- Institutes and centers
  - Center of Evaluation for Global Action (CEGA)
    - http://cega.berkeley.edu/
  - Poverty Action Lab (PAL)
    - http://www.povertyactionlab.org
  - Innovations for Poverty Action (IPA)
    - http://www.poverty-action.org/
- People on Campus
  - Me, Tap
  - Economics, Agricultural & Resource Economics
  - Haas school of business
  - School of Public Health

Final thoughts

- Hopefully, you better understand
  - What is impact evaluation
  - Why it's important
  - Why randomization is the easiest option
    - How to think about randomization
  - How you might do Impact Evaluation without randomization
    - Probably would want to consult an econometrician

General questions?