Capacity Building workshop on Impact Evaluation of Employment Programs

Choosing an Impact Evaluation method

Celine Ferre, Gdańsk, February 23, 2017
IE Methods Toolbox

Choose Your Method

- Randomized Assignment
- Discontinuity Design
- Difference-in-Differences
- Matching + Diff-in-Diff
A. LET’S RECAP… FOUR KEY ELEMENTS FOR AN IE
1. Clear understanding of the intervention
   
   Research question: “what are we evaluating?”

2. Well defined outcomes (impacts)
   
   What is/are the goal(s) of our policy?

3. Credible identification strategy/ies (definition of the counterfactuals)
   
   What methodology should we choose?

4. Reliable data
<table>
<thead>
<tr>
<th>METHODOLOGY</th>
<th>DESCRIPTION</th>
<th>WHO IS IN THE COMPARISON GROUP?</th>
<th>THE METHODOLOGY IS ONLY VALID IF...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before-and-after</td>
<td>Measure how program participants improved (or changed) over time.</td>
<td>Program participants themselves—before participating in the program.</td>
<td>The program was the only factor influencing changes in the outcome over time. If the program did not exist, outcomes would be the same before and after the study period.</td>
</tr>
<tr>
<td>Enrolled vs not-enrolled</td>
<td>Measure the difference between program participants and non-participants after the program is completed.</td>
<td>Individuals who did not participate in the program (for any reason), but for whom data were collected after the program.</td>
<td>Non-participants and participants were equally likely to enter the program before it started. Non-participants are identical to participants, except they did not participate in the program.</td>
</tr>
<tr>
<td>Randomized Control Trial (RCT)</td>
<td>Random assignment (e.g. a coin toss or random number generator) determines who may participate in the program so that those assigned to participate in the program are, on average, the same as those who are not, in both observable and unobservable ways. Since the participants and nonparticipants are comparable, except that one group received the program, any differences in outcomes result from the causal effect of the program.</td>
<td>Participants who are randomly assigned to not participate in the program. This is often called the “control” group.</td>
<td>Randomization “worked” and the two groups are statistically identical (on observed and unobserved factors). The effects of the treatment do not spill over to the control group. Any behavioral changes are driven by the program—not by the evaluation itself, or by the fact that the participants or non-participants are being studied. If outcome data are missing, data for the same types of individuals are missing from both the control and treatment groups.</td>
</tr>
<tr>
<td>METHODOLOGY</td>
<td>DESCRIPTION</td>
<td>WHO IS IN THE COMPARISON GROUP?</td>
<td>THE METHODOLOGY IS ONLY VALID IF...</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Regression Discontinuity Design (RDD)</strong></td>
<td>Individuals are ranked or assigned a score based on specific, measurable criteria. A cutoff determines whether an individual is eligible to participate in the program. Participants who are just above the cutoff are compared to non-participants who are just below the cutoff.</td>
<td>Individuals who are close to the cutoff, but fall on the “wrong” side of that cutoff, and therefore do not get the program.</td>
<td>After adjusting for the eligibility criteria (and other observed characteristics), the individuals directly below and directly above the cut-off score are statistically identical. The cutoff criteria must have been strictly adhered to. The cutoff must not have been manipulated to ensure that certain individuals qualify for the program.</td>
</tr>
<tr>
<td><strong>Difference-in-differences (DiD)</strong></td>
<td>Measure the before-and-after change in outcomes for the program participants, then subtract the before-and-after change in outcomes of the non-participants to find the relative change in outcomes for program participants.</td>
<td>Individuals who did not participate in the program (for any reason), but for whom data were collected both before and after the program.</td>
<td>If the program had not existed, the participants and non-participants would have experienced identical trajectories during the study period. Any differences in characteristics between the treatment and control group do not have more or less of an effect over time on outcomes.</td>
</tr>
<tr>
<td><strong>Propensity Score Matching (PSM)</strong></td>
<td>Individuals who received a program are compared to similar individuals who did not receive it.</td>
<td>Propensity score matching: For each participant, a non-participant with the same likelihood of participating, as predicted by known characteristics such as age, gender, and occupation.</td>
<td>The characteristics that were not included (because they are unobservable or have not been measured) either do not affect outcomes or do not differ between participants and non-participants.</td>
</tr>
</tbody>
</table>
B. PRACTICAL CONSIDERATIONS WHEN IMPLEMENTING AN IE
How do I choose the identification strategy and methodology?

Use opportunities to generate good comparison groups and ensure baseline data is collected.

3 questions help determine which method is appropriate for a given program:

- **Money**: Does the program have sufficient resources to achieve scale and reach full coverage of all eligible beneficiaries?

- **Timing**: How are potential beneficiaries enrolled in the program – all at once or in phases over time?

- **Targeting Rules**: Who is eligible for program benefits?
  
  Is the program targeted based on an eligibility cut-off or is it available to everyone?

Can we rank beneficiaries based on eligibility?
## Choosing your IE method(s)

<table>
<thead>
<tr>
<th>Money</th>
<th>Excess demand</th>
<th>No Excess demand</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Targeting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Timing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Phased Roll-out</strong></td>
<td>+ Randomized assignment + RDD</td>
<td>+ Randomized assignment + Randomized promotion + DD with matching + Randomized assignment + RDD</td>
</tr>
<tr>
<td><strong>Immediate Roll-out</strong></td>
<td>+ Randomized assignment + RDD</td>
<td>+ Randomized assignment + Randomized Promotion + DD with matching + RDD</td>
</tr>
</tbody>
</table>
Practical considerations (1)

• Impact evaluation is not for every intervention
  – Be selective
  – Be opportunistic

• The “gold standard” is plausible causality, not a single impact evaluation method
  – Recognize constraints
  – Be flexible, be creative

• Start early, work IE into the design of the program

• Think hard about benefits (what impacts to measure)
  – Link to project objectives
  – Careful choice of indicators
  – Understand time frame for outcomes to materialize
  – Identify logical axes of disaggregation (e.g. income groups, gender) and plan sample accordingly
Practical considerations (2)

• Monitor implementation of program – policy does not always equate to reality (know what you are evaluating) The same holds true for data collection
  – The task for implementing IE does not end with a sound design...

• Mix methods – qualitative and quantitative
  – Qualitative data provides information on the actual mechanism that caused the impacts
  – It may also provide intuition, new questions, and anecdotal stories that will enhance the final evaluation

• Watch for contamination of the treated and comparison groups
  – To the extent possible, bulletproof the control and treated groups
  – Stay on top of the implementation so we adjust the unforeseen events with sound solutions as quickly as possible
Practical considerations (3)

• It is important to work on Monitoring and Evaluation
  – This may link the IE with the ongoing efforts for monitoring the project
  – This generates more ownership of the IE by the project team

• It may be worth considering the implementation of Information Systems: targeting, program implementation, and evaluation

• Discuss your design with other IE colleagues
  – We learn from each other and you may get new ideas
  – This helps with later dissemination

• Work together with local partners
  – This may build local capacity for future evaluations
  – Validates the design and results
  – Having people in the ground helps to preserve the design
  – They tend to stand for change in governments
The Danish Case: systematic RCTs

- The Danish National Labor Market Board: Program for obtaining empirical evidence on effectiveness of ALMPs
  - RCTs – on average 2 per year
  - Knowledge-bank – database of empirical studies conducted worldwide on ALMP effectiveness
  - Works actively with research community, providing data, participating in conferences, interactive debates, etc.

- Status so far (2011)
  - 8 RCTs are completed
  - 2 ongoing
  - 1-2 in planning process
  - 30 small-scale ‘pilot projects’ planned

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Content</th>
<th>Region</th>
<th>Jobcentres</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Group meeting each week</td>
<td>Northern Jutland</td>
<td>Frederikshavn, Bronderslev, Hjørring</td>
</tr>
<tr>
<td>B</td>
<td>Early activation (after 13 weeks)</td>
<td>Mid Jutland</td>
<td>Aarhus</td>
</tr>
<tr>
<td>C</td>
<td>Early activation and group meeting each week</td>
<td>Southern Denmark</td>
<td>Esbjerg, Vejle</td>
</tr>
<tr>
<td>D</td>
<td>Individual meeting w. case worker every other week</td>
<td>Copenhagen &amp; Sealand</td>
<td>Holbæk, Roskilde, Gribskov, Ishøj, Vallensbæk, Vordingborg</td>
</tr>
</tbody>
</table>
C. USUAL CONCERNS
Some of the usual *ethical* concerns

- We can not run “experiments” on development issues
- We cannot leave people aside for the sake of the IE

**but...**
- Experimenting is part of the way to learn what is working (and what is not)
- We can not intervene with everybody anyway under budget/logistical constraints ➔ the evaluation may be a fair assignation

**Lessons Learned:**
- Work with counterparts at the beginning on identifying and addressing their concerns (mix political and research constraints)
- Be clear and explain everything
- Offer the evaluation as a solution rather than an extra layer for complications
Some of the usual political concerns

• There is no interest in showing (potential) bad news
• Long durations of the evaluations do not reconcile with political timelines

but...

– It is worse to do something bad and to hide it
– IEs may be designed as a tool to find “areas of improvement”
– The evaluation can be designed to pilot different options
– A good design can go beyond a political cycle

Lessons Learned:

– Understand the political concerns in order to design accordingly
– Work in phase
– Evaluate a subset of the program (in terms of geographic coverage)
– Show results soon (even with some limitations in the analysis) as it keeps politicians interested and engaged with the IE
Some of the usual **technical** concerns

- We already “know” what works... there is no need for evaluation
- The project is already complicated and we don’t want to add more complexity
- The concept of the project is already agreed upon
- The evaluations are too expensive, we cannot afford it

**but...**
- It may be complicated, but if we are not evaluating we are not learning
- Correlation and causality are not the same
- In most of the cases, the so called “agreements” are just basic features in the project’s concept

**Lessons Learned:**
- Projects in the preparation stage are better candidates than those under implementation
- Be pragmatic
Remember

The objective of impact evaluation is to estimate the **causal** effect or **impact** of a program on outcomes of interest.
Remember

To estimate impact, we need to estimate the counterfactual.

• what would have happened in the absence of the program and
• use comparison or control groups.
Remember

Choose the best evaluation method that is feasible in the program’s operational context.