Impact Evaluation for Safety Nets

Patrick Premand, Social Protection & Jobs GP
Safety Net Core Course, May 2018

Outline

From M&E to Impact Evaluation

The Main Concepts of Impact Evaluation

Choosing the best IE design for your project

Randomization as an operational tool

Interpreting IE results – Colin
From M&E to Impact Evaluation
The Results Chain in a Typical Program

Financial, human, and other resources mobilized to support activities.

Actions taken or work performed to convert inputs into specific outputs.

Project deliverables within the control of implementing agency SUPPLY SIDE.

Use of outputs by beneficiaries and stakeholders outside the control of implementing agency DEMAND SIDE.

Changes in outcomes that have multiple drivers.

Focus of traditional M&E

Focus of Impact Evaluation

Results-based management
Evaluations

A systematic, objective assessment of an on-going or completed project, program, or policy, its design, implementation and/or results, asking

- **Descriptive Questions** to seek to determine what is taking place and describe aspect of a process.

- **Normative Questions** to compare what is taking place to what should be taking place. *(PROCESS EVALUATION)*

- **Cause-and-Effect Questions** to examine outcomes and assess what difference the intervention makes in outcomes *(IMPACT EVALUATION)*
Impact Evaluation is not for every project

- Evaluate impact selectively, when project is:
  - Innovative
  - Replicable, scalable, or implemented at scale
  - Strategically relevant (e.g. large budget)
  - Evaluation will fill knowledge gap
  - Substantial policy impact

- Impact Evaluation can focus on selective innovations within projects
  - Beyond ‘does my program work’?
  - Towards ‘which design is more effective?’
Public Works (THIMO) in Cote d’Ivoire

- Emergency Youth Employment and Skills Development Project - PEJEDEPC (US $100 million)

- Public Works Component
  - 12,500 youths (18-30 years old) by 2015
  - Daily wage rate of CFA 2,500 for 6 months

- Graduation elements in public works:
  - Entrepreneurship training to help entry into self-employment
  - Job search training to help youth transition into wage jobs

- (Also: payment into bank accounts, basic life skills training)
Impact Evaluation Questions: Public Works in Cote d’Ivoire

**Basic Question**
What is the impact of participation in the public works program on youth employment and earnings?

**Design Question (1)**
Does the entrepreneurship training facilitate more productive self-employment after exit from the program?

**Design Question (2)**
Does the job search training facilitate insertion into wage jobs after exit from the program?

Results: [here](#) (French) or [here](#) (English)
The Main Concepts of Impact Evaluation
Impact Evaluation needs to be distinguished from other “evaluations”

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

Estimate the causal effect (impact) of intervention (P) on outcome (Y).

(P) = Program or Treatment
(Y) = Outcome Indicator, Measure of Success

Example: What is the effect of a cash transfer program (P) on Household Consumption (Y)?
Solution
Estimate what *would* have happened to outcomes \((Y)\) in the absence of the program \((P)\).
We call this the **Counterfactual**.

The key to a good impact evaluation is a valid estimate of the **counterfactual**!

Counterfactual is estimated based on **comparison groups**
Communicating complex concepts in 3 slides... Example: What is the Impact of...

giving Ruslan additional money on Ruslan’s consumption (P)?
The Perfect Clone

Ruslan

6 candies

Ruslan’s Clone

4 candies

IMPACT = 6 - 4 = 2 Candies
In reality, use statistics

Impact = 6 - 4 = 2 candies
Choosing the best IE design for your project
Finding good comparison groups

We want to find clones for the Alis in our programs.

The treatment and comparison groups should
- have identical characteristics except for benefiting from the intervention.

With a good comparison group, the only reason for different outcomes between treatments and controls is the intervention (P).
Two false counterfactuals to avoid

Before vs After

**Compare:** Same individuals Before and After they receive P.

**Problem:** Other things may have happened over time.

Enrolled vs Not Enrolled

**Compare:** Group of individuals Enrolled in a program with group that chooses not to enroll.

**Problem:** Selection Bias. We don’t know why they are not enrolled.

Both counterfactuals lead to biased estimates of the counterfactual and the impact.
The conversation needs to start early

"Retrospective Evaluation is necessary when we have to work with a program that has already been roll-out and existing data. Rarely feasible: baseline data? Information on targeting?"

"In Prospective Evaluation, the evaluation is designed in parallel with the program (and targeting decisions). The way to go: ensure baseline data is collected, and comparison group exists."
Where do good Comparison Groups come from?

- The rules of program operation determine the evaluation strategy.
- We can almost always find a valid comparison group if:
  - the operational rules for selecting beneficiaries are equitable, transparent and accountable;
  - the evaluation is designed prospectively.

- Evaluation design and program design go hand-in-hand.
5 methods in IE Toolbox take different approaches to generate comparison groups and estimate the counterfactual:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Randomized Assignment</td>
</tr>
<tr>
<td>2</td>
<td>Randomized Promotion</td>
</tr>
<tr>
<td>3</td>
<td>Regression Discontinuity Design</td>
</tr>
<tr>
<td>4</td>
<td>Difference-in-Differences</td>
</tr>
<tr>
<td>5</td>
<td>Matching</td>
</tr>
</tbody>
</table>
Choosing an IE design for your program

- Design IE prospectively to generate good comparison groups and collect baseline data.
- 3 operational questions to 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?
  
  **Targeting Rules:** Who is eligible for program benefits? Is the program targeted based on an eligibility cut-off or is it available to everyone?
  
  **Timing:** How are potential beneficiaries enrolled in the program – all at once or in phases over time?
## 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>Targeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phased Roll-out</td>
<td>Targeted</td>
<td>Universal</td>
</tr>
<tr>
<td>Immediate Roll-out</td>
<td>1 Randomized assignment 4 RDD</td>
<td>1 Randomized assignment 2 Randomized promotion 3 DD with 5 Matching</td>
</tr>
<tr>
<td></td>
<td>1 Randomized Assignment 4 RDD</td>
<td>1 Randomized Assignment 2 Randomized Promotion 3 DD with 5 Matching</td>
</tr>
</tbody>
</table>

- **Phased Roll-out**: Suitable for scenarios with excess demand where targeting and timing can be carefully controlled. Phased roll-out allows for gradual implementation, making it easier to manage and adapt to feedback.
- **Immediate Roll-out**: Ideal for situations where immediate action is necessary. Immediate roll-out can be implemented swiftly but may require careful planning to avoid overwhelming beneficiaries or stakeholders.

### Notes
- **RDD**: Randomized Design Disorder
- **DD**: Difference-in-Differences
- **Matching**: Techniques used to control for confounding variables.
# Choosing the IE method in Niger

<table>
<thead>
<tr>
<th>Money</th>
<th>Excess demand</th>
<th>No Excess demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phased Roll-out</th>
<th>Targeted</th>
<th>Universal</th>
<th>Targeted</th>
<th>Universal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Randomized assignment</td>
<td>1 Randomized assignment</td>
<td>1 Randomized Assignment</td>
<td>1 Randomized assignment to phases</td>
<td>1 Randomized assignment to phases</td>
</tr>
<tr>
<td>4 RDD</td>
<td>2 Randomized promotion</td>
<td>2 Randomized promotion</td>
<td>2 Randomized Promotion to early take-up</td>
<td>2 Randomized Promotion to early take-up</td>
</tr>
<tr>
<td>3 DD with 5 Matching</td>
<td>3 DD with 5 Matching</td>
<td>4 RDD</td>
<td>3 DD with 5 matching</td>
<td>4 RDD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Immediate Roll-out</th>
<th>Targeted</th>
<th>Universal</th>
<th>Targeted</th>
<th>Universal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Randomized Assignment</td>
<td>1 Randomized Assignment</td>
<td>1 Randomized Assignment</td>
<td>1 Randomized assignment to phases</td>
<td>1 Randomized assignment to phases</td>
</tr>
<tr>
<td>4 RDD</td>
<td>2 Randomized Promotion</td>
<td>2 Randomized Promotion</td>
<td>2 Randomized Promotion</td>
<td>2 Randomized Promotion</td>
</tr>
<tr>
<td>3 DD with 5 Matching</td>
<td>3 DD with 5 Matching</td>
<td>5 Matching</td>
<td>5 matching</td>
<td>5 Matching</td>
</tr>
</tbody>
</table>

If less than full Take-up: 2 Randomized Promotion 3 DD with 5 matching
Randomization as an operational tool
Randomization is not only an evaluation tool. In Côte d’Ivoire...

- Public works were introduced as an instrument right after a period of conflict.
- How to chose between the large number of youth interested to participate?
- Public lottery was deemed the most transparent, and least controversial approach.
- Project decided to use randomization/lottery for purely operational reasons.
- The impact evaluation came later and just built on this.
Randomization can help with transparency

Project staff in Niger: “Now political authorities cannot interfere with the village selection. All the village chiefs were present and signed that they agreed with the procedure before we did the selection. No one can complain to us and try to change the result.”

Beneficiaries in Nicaragua: “At least this time we know why we were not chosen for the program. Usually decisions are made and we don’t know why our village cannot participate.”
Randomization

- Randomization can be used only in certain contexts.
- BUT excess demand happens for most programs
  - Even after applying all existing targeting criteria, not everyone can be served
  - Randomization is fair, transparent and ethical way to assign benefits to equally deserving populations
  - Provides equal chance of participation among equally deserving units.
- Randomization is the Gold Standard for impact evaluation
  - Most robust method
  - But also the most simple, and the cheapest.
- Multiple ways to perform randomization
Randomization to answer basic IE questions

1. Population

2. Evaluation sample

3. Randomize treatment

External Validity

Internal Validity

= Ineligible

= Eligible
Randomized to answer IE “design” questions

1. Population

2. Evaluation Sample


Not eligible

Eligible
Randomized Assignment

In Randomized Assignment, large enough samples, produces 2 statistically equivalent groups.

We have identified the perfect clone.

Feasible for prospective evaluations with over-subscription/excess demand.

Most pilots and new programs fall into this category.

Consider evaluating relative effectiveness of alternative program design options.
Key Messages
The objective of impact evaluation is to estimate the **causal** effect or **impact** of a program on outcomes of interest.
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.
We have a **toolbox** with **5 methods** to identify good comparison groups.
Choose the best evaluation method that is feasible in the program’s operational context.
Thank you!

Reference also: available in Spanish, French and Portuguese

www.worldbank.org/ieinpractice
Operational aspects
What to Evaluate?

Efficacy Studies are carried out in a specific setting to test a “model” implemented in best-possible way.
(e.g. Pilots for proof of concept)

Effectiveness Studies, provide evidence from interventions taking place under normal circumstances
(e.g. Scalable National Programs)
Who does the Impact Evaluation?

- Critical to start discussing IE early:
  - Clarify role of different types of “evaluations”
  - Large potential value-added of IE… but it is an investment

- Essential to design the evaluation with operational team
  - Framing of evaluation question
  - Program design and IE design go together.
  - Implementing of IE requires close coordination with project implementation

- IE best as seen of collaboration between implementers and evaluators
  - Quality/Validity of design is what makes results legitimate
  - Consider which components to outsource
Design and implementation of IE in Niger?

• Broad outline of Impact Evaluation in project design document
• After effectiveness, 6-months preparation phase (typically longer)
  • Multiple workshops with local stakeholders to finalize design
  • Participation of project team in a regional impact evaluation workshop
• Randomization performed by project (through geographical targeting process)
• Baseline data collection contracted to National Statistical Agency (June 2012)
  • Oversight by WB and project teams.
• Next steps:
  • Focus on ensuring MIS colleges comprehensive data on project implementation
  • Qualitative (process) evaluation to gather information on quality of implementation, as well as beneficiaries’ perceptions
  • Follow-up data collection (December 2014)
Benefits of Impact Evaluation

• IE is the only way to know if a program is effective.
  • Clear value-added, but needs to be used selectively.

• Doing IEs can improve project implementation
  • The process of doing IE can change the way we work (“Science of Delivery”)

• Learning through the process in Niger:
  • Learning quality control in baseline helped with PMT survey
  • Randomization helped with transparency
  • Solid baseline helps for profiling of beneficiaries
  • Allows analysis of targeting efficiency
  • Complementary qualitative evaluation helps improve project implementation
Impact Evaluation, MIS, and Process evaluation are complementary

- Monitoring and Process Evaluation data is key for the evaluation
  - MIS: Lists of beneficiaries, distribution of benefits
  - Targeting data
  - Process evaluation

- Common issues:
  - For the IE design, can you do targeting (e.g. PMT) in comparison group?
  - How to set up unique identifiers across all sources of data

- Data on project implementation is essential to interpret IE result. Does the project work/not work:
  - because of the intervention model,
  - or because of implementation of the intervention model?