Randomised Evaluation in Action

Sonya Krutikova (IFS Associate Director, EDePo Director)
Randomised Evaluation - recap

• We want to base ECCE policy on evidence of what works

• Policy effectiveness evidence base needs to be built up in MENA countries (and worldwide)

• First-best are randomised evaluations as they provide the most credible *counterfactual*
Questions that can be answered by a randomized evaluation

- Was the programme effective and how effective?
Questions that can be answered by a randomized evaluation

• Was the programme effective and how effective?
• Why did it work or not work?
Questions that can be answered by a randomized evaluation

- Was the programme effective and how effective?
- Why did it work or not work?
- Were there unintended side-effects?
Questions that can be answered by a randomized evaluation

• Was the programme effective and how effective?
• Why did it work or not work?
• Were there unintended side-effects?
• Who benefited the most and was anyone harmed?
Questions that can be answered by a randomized evaluation

- Was the programme effective and how effective?
- Why did it work or not work?
- Were there unintended side-effects?
- Who benefited the most and was anyone harmed?
- What lessons can be applied to other contexts or if the programme was scaled up?
Questions that can be answered by a randomized evaluation

• Was the programme effective and how effective?
• Why did it work or not work?
• Were there unintended side-effects?
• Who benefited the most and was anyone harmed?
• What lessons can be applied to other contexts or if the programme was scaled up?
• How cost-effective was it & how does it compare to other programmes targeting similar goals?
The main impact evaluation challenge

...is to answer the question:

“How would those who received the programme or benefited from a specific policy have fared otherwise?”
The main impact evaluation challenge…is to answer the question:

“How would those who received the programme or benefited from a specific policy have fared otherwise?”

But impossible to observe the same person in two mutually exclusive states.
The main impact evaluation challenge

...is to answer the question:

“How would those who received the programme or benefited from a specific policy have fared otherwise?”

But impossible to observe the same person in two mutually exclusive states

So who can we compare those who receive the programme to?
The solution in randomised evaluation

• In randomised evaluation the comparison group is created through random assignment.

• This ensures that those who receive the programme and those who do not are representative of the larger group that both come from and are statistically equivalent to each other.

• Other methods of constructing the comparison group in other ways has much more risk of producing misleading results.
But not always!

*Not all policies/programmes are suitable for randomised evaluation*

- Evaluation macro policies
- Programme is at very early stages – still evolving and changing
- Unethical to deny the control group the intervention
- Too difficult to control factors which affect integrity of the experiment
- Sample size is too small
But not always!

*Not all policies/programmes are suitable for randomised evaluation*

- Evaluation macro policies
- Programme is at very early stages – still evolving and changing
- Unethical to deny the control group the intervention
- Too difficult to control factors which affect integrity of the experiment
- Sample size is too small

© Institute for Fiscal Studies
Randomised evaluation if life were simple

Step 1: We choose a group of interest e.g. children age 3-5 who do not go to pre-school.
Randomised evaluation if life were simple

Step 1: We choose a group of interest e.g. children age 3-5 who do not go to pre-school.

Step 2: We flip a coin/throw a die/draw names out of a hat to allocate the children to 2 groups
Randomised evaluation if life were simple

**Step 1:** We choose a group of interest e.g. children age 3-5 who do not go to pre-school.

**Step 2:** We flip a coin/throw a die/draw names out of a hat to allocate the children to 2 groups

**Step 3:** one group gets offered places at pre-school, the other does not
Randomised evaluation if life were simple

**Step 1:** We choose a group of interest e.g. children age 3-5 who do not go to pre-school.

**Step 2:** We flip a coin/throw a die/draw names out of a hat to allocate the children to 2 groups

**Step 3:** one group gets offered places at pre-school, the other does not

**Step 4:** a year later we compare measures of cognitive and socio-emotional development among the groups that did and did not receive the pre-school place to evaluate the impact of pre-school on child development.
Sounds good but life is not simple

- How doable is that in context of on-the-ground realities of policy development and implementation.
Sounds good but life is not simple

• How doable is that in context of on-the-ground realities of policy development and implementation.

More doable than might seem at first....but a lot of the work is upfront (unlike picking up an existing data-set)

→ Design is key
Randomised evaluation in practice: conceptually

- What do we want to learn from the evaluation?
- What are the relevant questions?
- What outcomes are expected?
- How can they be measured?
Randomised evaluation in practice: design

• What is the appropriate unit of randomisation?
• What is the appropriate method of randomisation?
• What are the main threats to the design?
• How will we implement the randomisation?
• What is the sample size?
Briefly - ECCE in Colombia

• 2011 national early education strategy “From Zero to Forever” DCAS launched in response to evidence on importance of pre-primary education for growth

• Aim = to deliver high quality integrated ECD services to disadvantaged children

• Many programmes exist and are targeted at different socio-economic groups. Eligibility based on asset score and parental labour force participation.

• Many new programmes implemented in collaboration with researchers in ways that have allowed to build in randomised evaluation. Increasingly new policies are taking into account findings from these.
Randomised evaluation in practice: design

- What is the appropriate unit of randomisation?
- What is the appropriate method of randomisation?
- What are the main threats to the design?
- How will we implement the randomisation?
- What is the sample size?
Randomisation unit: not necessarily unit of analysis

Andrew et al, *Evaluation of Hogares Infantiles (child-care centres for low SES children of working parents)*

2 arm trial of structural and process quality improvements to pre-school.

Interested in child outcomes but would not want to randomise at child level:

1. Unit of analysis → child
2. Unit targeted for treatment → class
3. At what levels might there be spillovers (will come back to spillovers) → pre-school, educational authority
4. What’s a feasible sample-size → not feasible to randomise at educational authority level but could (and did) randomise at pre-school level (more about this at the end)
Randomised evaluation in practice: design

- What is the appropriate unit of randomisation?
- What is the appropriate method of randomisation?
- What are the main threats to the design?
- How will we implement the randomisation?
- What is the sample size?
Randomisation designs – exploiting on-the-ground opportunities

• May seem like randomisation requires a daunting “subversion” of the policy implementation process

• I hope to show that often:
  – There is more flexibility in the design than might seem at first
  – Designs can be built around specific features/constraints of implementation of a programme
  – Designs can even help improve implementation process (e.g. making programme assignment criteria more fair)
Methods of randomisation: exploiting on-the-ground opportunities

- **Limited political/funder support:** it may be necessary to strengthen the case for the programme / insulate it against political change (e.g. PROGRESA withstood change of govt thanks to randomized evaluation evidence).

- **Limited resources:** there may not be enough money, staff, skills, etc to include everyone who would be eligible

- **Limited capacity:** it may not be possible to roll out the programme to everyone from the outset

- **Ensure minimum level of take-up/encourage take-up** among priority group without fully excluding others
Simple Lottery
Randomly sample from area of interest

Randomly assign to treatment and control

Randomly sample from both treatment and control

Simple lottery
Simple lottery: example

Attanasio et al. *Effects of an early stimulation curriculum and a nutritional intervention in a national public parenting program in Colombia on child development and growth: a cluster-randomized controlled trial*, 2016

Evaluate effect of early stimulation and nutritional intervention implemented through public parenting support services for families living in rural Colombia facilitated by women in the community through home visits.
Evaluation Design

- Cluster randomised trial due to concerns about spillovers. Randomised at municipality level
- Started with 135 eligible municipalities
- Excluded 39 for logistical reasons
- Of the remaining 96, 49 randomly assigned to treatment where FAMI was implemented & 47 to control
- Within these sampled a total of 1460 eligible children under the age of 1
- Found positive effects on cognitive development, reduction in malnutrition, improved parenting practices at home
# Simple lottery: pros and cons

<table>
<thead>
<tr>
<th>Design</th>
<th>Most useful when</th>
<th>Advantages</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Lottery</td>
<td>• Limited resources / programme oversubscribed</td>
<td>• Easy to understand &amp; implement</td>
<td>• Control group may not cooperate</td>
</tr>
<tr>
<td></td>
<td>• Proof of concept</td>
<td>• Can be done as a public lottery</td>
<td>• Differential attrition</td>
</tr>
</tbody>
</table>
Phase-In
Phase 0:  
No one treated yet  

All control
Phase 1:
1/4 treated
3/4 control
Phase 2:
2/4 treated
2/4 control
Phase 3:
3/4 treated
1/4th control
Phase 4:
All treated (experiment over)

- *Hogares Communitarios* provide semi-formal community based care for children under 5. It’s an important type of child-care for children from low-SES homes. More than half of ECE provision for these children was through HC’s since the 1980’s

- Vocational educational programme introduced in 2007 to offer a degree to the care-providers motivated by findings from randomized evaluations of HC’s of very low levels of education and knowledge of infant and child development and low levels of quality of care provision.

- Govt developed vocational education degree over 2 semesters, 2640 hours of instruction, offered at no cost

- Evaluation exploits gradual expansion in capital (Bogota) to evaluate impacts. Programme offered sequentially in 3 phases by neighbourhoods with neighbourhoods selected randomly.

- Programme found to improve both quality of care and child health and cognitive development
Phase in: example

Control group: Randomly selected 80. 73 out of 80 interviewed.
Control group initiates ECDC program

Call for participation issued in 3 neighborhoods (3) 198 MC expressed interest

First cohort:
72 MC enter the program (1)

First cohort:
43 MC graduate from the program

Second cohort:
78 MC enter the program (2)

Second cohort:
54 MC graduate from the program

Treatment group:
Randomly selected 80. 67 out of 80 interviewed. 26 first cohort 41 second cohort

Data collection 140 HCBs

# Phase-in: pros and cons

<table>
<thead>
<tr>
<th>Design</th>
<th>Most useful when</th>
<th>Advantages</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase-in</td>
<td>• Everyone will receive treatment <em>eventually</em> but not all at the same time</td>
<td>• Easy to understand &amp; explain</td>
<td>• Control group may change behaviour in anticipation of treatment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Control group cooperates as expected to benefit</td>
<td>• Limits opportunities to measure long-term impact</td>
</tr>
</tbody>
</table>
Encouragement design

Sometimes it’s just not possible to not offer a programme to everyone e.g. when there are enough resources and a pretty good hunch that it is beneficial / govt wants to implement programme fast

Can, in some cases, still make quite easy tweaks to implement in a way that is amenable to robust evaluation if:

• Expect fairly low take-up

• BUT there are practical and inexpensive ways to effectively encourage take-up

Then:

• Can randomize the encouragement and compare those who received the encouragement to those who didn’t
Encouragement design

- Encourage
- Do not encourage

- Participated
- Did not participate

compare encouraged to not encouraged

do not compare participants to non-participants
Encouragement design

- Encourage
- Do not encourage
- Participated
- Did not participate

Compare encouraged to not encouraged
Do not compare participants to non-participants

These must be correlated
Encouragement design: example

Attanasio et al, *The Effects of the Transition from Home-based Childcare to Childcare Centers on Children’s Health and Development in Colombia*, 2016

• A key pillar of the 2011 national early childhood strategy is provision of integrated / comprehensive early childhood services which concurrently provide childcare, health & nutrition, education, recreation etc

• HC’s were classed as inadequate for this & children were offered opportunity to transfer to large childcare centres (*Centros de Desarrollo Infantil* CDI) which have age-specific groups, better facilities, more staff, more structure and adherence to guidelines. Annual cost of child at CDI=USD 1600 compared to less than a quarter of that at CDI.

• Use clustered-randomized trail to evaluate impacts of the offer made to move from HC to CDI.
Encouragement design: example

Chose 14 cities and 15 CDI’s up and running by the time of the study.

For each CDI identified 28 surrounding HC’s – took 20 of these and randomly allocated 15 to treatment and 5 to control

Treated HC’s received offer to transfer, control did not. MC’s were meant to encourage parents
## Encouragement design: pros and cons

<table>
<thead>
<tr>
<th>Design</th>
<th>Most useful when</th>
<th>Advantages</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encouragement</td>
<td>• Programme has to be available to all at the same time</td>
<td>• Can randomise at individual level even when programme is not administered at that level</td>
<td>• Only measuring impact on those who responded to incentive</td>
</tr>
<tr>
<td></td>
<td>• Take-up is low but can be improved with simple incentives</td>
<td>• Want to look at impacts on a particular sub-group</td>
<td>• Adds complexity - need something that’s enough of an incentive but doesn’t have a direct effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• More sample size issues</td>
</tr>
</tbody>
</table>
Randomised evaluation in practice: design

- What is the appropriate unit of randomisation?
- What is the appropriate method of randomisation?
- What are the main threats to the design?
- How will we implement the randomisation?
- What is the sample size?
Spillovers & cross-overs

- Happen when individuals in control group are somehow affected by the treatment:
  - Interaction with the treated
  - Joining treatment group

- Can reduce risk by anticipating and adjusting unit of randomisation
  - Thinking through potential sources of spillovers
  - Spacing randomisation units

- Can be quite tricky when balancing multiple requirements and constraints
Example: Randomised evaluation of pre-school intervention in Ghana
Attrition

• Happens when we aren’t able to collect data on the full sample of individuals for who were selected as part of the original sample

• Possible reasons:
  • Working with a highly mobile population
  • Participants withdraw

• Risks to evaluation:
  • Small sample problem making it harder to detect effects
  • Compromises study design if attrition is different across groups, e.g.:
    • Participants in control group get annoyed
    • Programme affects migration rates & no tracking

• Solutions:
  • Timing of survey rounds
  • Selecting appropriate design from the start
  • Raport with sample
  • (Ex-post rebalancing of the sample)
Randomised evaluation in practice: design

- What is the appropriate unit of randomisation?
- What is the appropriate method of randomisation?
- What are the main threats to the design?
- How will we implement the randomisation?
- What is the sample size?
Feasible Randomisation Mechanics

• **Simple Lottery:**
  • Have complete list of people/households/pre-schools....randomisation units
  • Write a computer programme which does the equivalent of drawing names out of a hat

• **Spot Randomisation:**
  • Programme entry has to be determined on a rolling basis
  • Implementers "flip a coin" / computerised randomisation
  • *Example:* Unemployed Danish youth assigned to government vs alternative programme when they come to the job-centre.
    – Case worker classifies youth as eligible for the alternative programme or not
    – If eligible, logs into a randomisation programme and is told whether to assign individual to treatment or control
    – Randomisation programme allocation ratio can be adjusted depending on relative demands of the two programmes
Randomised evaluation in practice: design

- What is the appropriate unit of randomisation?
- What is the appropriate method of randomisation?
- What are the main threats to the design?
- How will we implement the randomisation?
- **What is the sample size?**
Small samples

• Never have unlimited resources for evaluation so face many trade-offs e.g. sample size versus number of questions we can ask, treatment arms we can have

• Will talk more about how to decide on appropriate sample size this week but for now just honing in on a sample size issue relevant for randomisation methods.

• With "small" sample sizes, may want to _stratify_ the sample into sub-groups before randomization so that there is balanced across the groups

• (May also want to look at results by sub-group)
Stratification: example

Total of 670 HI’s in Study cities (Bogotá, Cali, Medellín, Barranquilla, Bello, Palmira, Itagüí and Soledad)

198 randomly selected & organised into geographically close groups of 3

40 groups (120 HI’s) selected based on having at least 15 children 18-36 months @ baseline

Random Assignment

Treatment 1 - HIM
Govt Improvements
40 HI’s
Total kids: 663

Treatment 2 - HIMFE
Govt + FE Improvements
40 HI’s
Total kids: 663

Control
40 HI
Total kids: 661

Total sample = 1987 kids 18-36 months
Concluding Remarks

• Building in a randomised evaluation does not have to totally disrupt the policy making/programme implementation process

• There is a lot of scope to think creatively and leverage existing opportunities

• However, key challenge is to anticipate potential challenges and feasibility considerations at the design stage. Very hard to fix later.

• Must start with a clear idea of what and why we want to know

• Then think of
  • what design would allow to get answers to these questions
  • what spillovers, cross-overs, sample size and attrition issues might arise
  • how design + data collection could pre-empt that
The End

Thank you!
Recap: basic RCT

Comprehensive evaluation framework (Rossi, Freeman & Lipsey)

• Needs Assessment
• Programme theory assessment
• Process evaluation
• Impact evaluation
• Cost-benefit, cost-effectiveness, and cost-comparison analysis
• Goals, outcomes and measurement
More methods of randomization

- Rotation
- Varying levels of treatment