ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Designs and Implementation issues in randomized evaluations: A toolkit

Esther Duflo

DARES Conference

22-23 May

Practical Issues about Randomized Evaluations

- Partners
- How to introduce randomization in a research Design?
  - Retain Rigor
  - …While being Practical
- Some issues on Analyzing data from experiments
- Design choices and their consequences
  - Clustered vs individual designs
  - Factorial Designs

Partners

- Many possible partners:
  - Governments
  - Local collectivities
  - NGOs
  - Private firms

How to Randomize: The key Constraints

- It must be operationally feasible
  - Compatible with program objectives
  - Field staff must be able to implement it on the field without major obstacles to their day-to-day work.
- It must be ethical and fair
- It must also be *perceived* as fair
- It must still give us the ability to compare two (or more!) randomly assigned groups.

Different types of Solutions

- The pure randomized pilot model:
  - Many examples yesterday (e.g. MTO, STAR, ERA, Bergen)
  - Most similar to a medical model
  - However in general, compliance is not enforced in the treatment group (we will see later how to deal with that).

Different types of Solutions

- The community based randomized cluster pilot:
  - Program is made available to a community/school etc.: All those eligible for the program in this community have access to it:
    - Yesterday's example: Israel experiments, Progresa, immunization experiment.

Different types of Solutions

- Randomized Phase-in:
  - Everybody in the study will receive the program at a point or another
  - But the program entry is phased in over a period of time
  - E.g. Deworming program in Kenya: 75 schools in total: 25 in year 1, 25 in year 2, 25 in year 3
  - The only issue with that method is that the phase in must be slow enough to leave time to see the effects (e.g. 6 months phase in for a micro-credit program will NOT WORK!)

Different types of Solutions

- Resource Constraints: Lottery
  - There are programs (often at the pilot stages) which simply do not have enough resources to treat everyone, and where randomizing is seen as a fair way to allocate the program
  - The task of the evaluation team is then to follow lottery "winners" and "losers" (often among applicants)
  - Examples:
    - ANPE/UNEDIC
    - Secondary school Voucher program in Colombia.
    - Magnet schools in the US, China.

Different types of Solutions

- Randomization “in the bubble”
  - Examples: E2C Paris, Microcredit in South Africa.
  - Often Program officers have three category of people:
    - The people they absolutely want to treat
    - The people they don’t want to have anything to do with
    - A set of people they would be happy to work with if they had more resources, but they are currently not serving (the “bubble”)
  - The randomization can exclude those in the first and second category and concentrate on the third one.
  - This gives us the impact on this group of people, which may no be representative of the population at large, but often represents a group of interest (since they would be targeted for an expansion of the program).

Different types of Solutions

- Encouragement Design:
  - How to evaluate a nationwide policy, or something everyone could do (like a flu shot).
  - There are policies that not everybody takes advantage of (e.g. not everyone gets a flu shot… not every unemployed person takes advantage of all the possibilities for training and help that they get).
  - An encouragement design: randomly select a group and gives them help/incentives/information to undertake an activity.

Encouragement Designs (cont.)

- Later on follow them up to :
  - Find out whether they undertook the activity (first stage)
  - Find out whether their outcome differ
- Evaluation design we proposed for the Adie microcredit program
- Advantage and disadvantages:
  - Very palatable to program officers
  - Won’t work if the encouragement design does not work very well!
  - Must be careful that the program does not affect the outcome directly (flu shots)

Analysis issues

- Attrition: Losing data
  - To avoid…
  - Particularly bad if attrition is differential
  - Can bound treatment effects under certain assumptions but if there is a lot of attrition the bounds will be wide.
  - Worth putting procedures in place to limit attrition:
    - Administrative data
    - Ways to keep in touch with subjects

Analysis issues

- Non compliance: some members of the treatment group did not get treated (and/or some member of the control group get treated
  - Many examples yesterday
  - Inherent to encouragement designs.
  - Less bad (to a point…)
  - KEY POINT: ALWAYS compare those initially assigned to the treatment and the control group. This is the intention to treat analysis
  - Two things to keep in mind:
    - Non compliance will affect power
    - By using Instrumental variable (divide ITT by difference in take up), it is possible to recover the effect of treatment on the treated (if one sided non-compliance) or of the treatment on the compliers, but we have to keep in mind this may or may not generalize/be a group of interest
    - Need to assume that randomization itself did not have any direct impact (some time an issue with encouragement designs).

Analysis issues

- Externalities:
  - Does being close to a treatment group affect the control group as well?
    - Contagion (deworming)
    - Peer effects (e.g. education effects)
    - Market equilibrium impacts (e.g. jeune chomeurs: do they take the place in the queue of someone).
  - Can underestimate Treatment effect when randomization is within group
  - Can overestimate when using IV if randomization is across group and there is imperfect compliance
  - Ideally when randomizing at two levels (community and then individual), can find out about them (e.g. information session in Harvard, jeune chomeurs).

Some designs issues

- Power of an experiment:
  - One minus the probability of being disappointed…
  - (more formally: 1-probability of not finding an effect of a given size at a given level of significance)
  - It is affected by:
    - Effect size
    - Sample size
    - Compliance
    - Clustered or not clustered design
  - Do not start underpowered design…
  - Do not let your self be carried away by the enthusiasm of your implementation partners

The Level of Randomization

- Some studies could be randomized at the individual level, or at a more aggregate level.
- Suppose for example you are planning to evaluate a microcredit program.
- What could be a possible individual level randomization design, and what could be a group level randomization design?
- What are the reasons to prefer one versus the other?

Clustered Design

Cluster randomized trials are experiments in which social units or clusters rather than individuals are randomly allocated to intervention groups

Examples:

PROGRESA	Village
Gender Reservations	Panchayats
Flipcharts, Deworming	school
Iron supplementation	Family

Reason for adopting cluster randomization

- Need to minimize or remove contamination
  - Example: In the deworming program, schools was chosen as the unit because worms are contagious
- Basic Feasibility considerations
  - Example: The PROGRESA program would not have been politically feasible if some families were introduced and not others.
  - Natural administrative unit (districts/blocks)
- Only natural choice
  - Example: Teacher Training

Impact of Clustering

- The outcomes for all the individuals within a unit may be correlated
  - All villagers are exposed to the same weather
  - All students share a schoolmaster
  - The program affect all students at the same time.
  - The member of a village interact with each other
- This needs to be taken into account when computing standard errors. First approximation, power increase with the number of units over which your randomized, NOT the number of subjects
- Randomizing one district in, one district out does is not a randomized evaluation!

Pro and cons or clustered designs

- Cons:
  - Sample size requirement
  - Evaluation is more geographically spread out
  - More individuals need to be involved
- Pro:
  - Contamination of control groups less likely
  - Feasibility in principle:
  - Feasibility in practice: risk of implementation slippage when individuals need to make decision in the field.
- A decision to be taken on a case by case basis. But remember to adjust the sample size accordingly!!!

Testing multiple hypotheses: factorial designs

- Factorial Design: Tests two intervention at the same time
- For example one may consider supplying SMI with access to credit and with business training

		Business training	
		YES	NO
Credit	YES	GROUP 1	GROUP 2
	NO	GROUP 3	GROUP 4

Uses of factorial design

- First use - test interactions between program:
  - Perhaps credit is useless without business training
  - Perhaps business training is useless without credit.
  - In this case study must be have enough power to distinguish all four cells from each other

Uses of factorial design

- Second use: Make evaluation more cost effective
  - Extensive data may be collected to evaluate the impact of credit
  - One might as well test the effect of business training as well - the intervention is very cheap. Both the credit intervention and the business training intervention will charge (**?**) costs.
- In this case, we may not need enough sample to separate each of the four groups from each other: we may be happy with enough sample to separate
- GROUP 1+ GROUP 2 vs GROUP3+GROUP 4 and GROUP 1+ GROUP 3 vs GROUP2+GROUP 4

Pros and Cons of factorial designs

- Pros:
  - Much more cost effective
  - Can inform us about what components of an intervention really works (counterexample: PROGRESA).

- Cons
  - When sample size are not large enough to estimate interaction: estimate effect of credit in a sample where half the sample receives training. May be a problem.
  - Implementation becomes tricky.