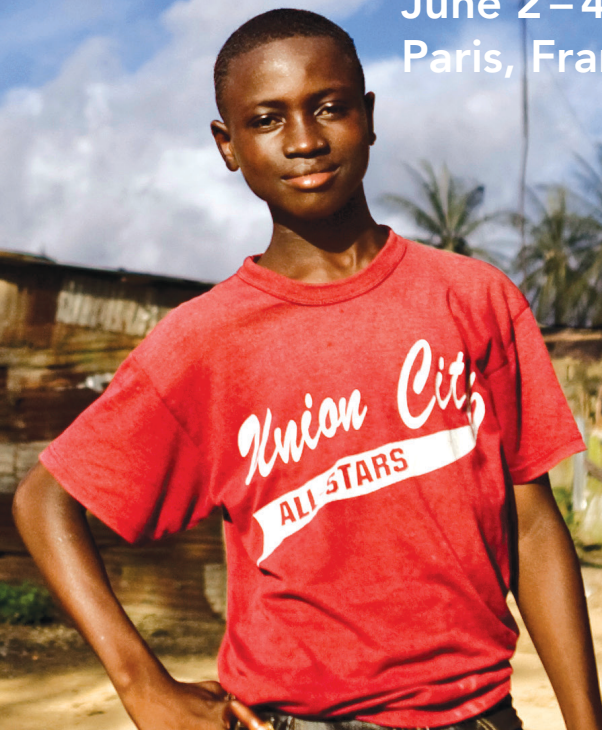J-PAL Post-Primary Education and Youth
Matchmaking Conference

June 2 – 4, 2014
Paris, France

ABDUL LATIF JAMEEL
Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Measuring Impacts
Principles Of Impact Evaluation And Randomized Trials

Bruno Crépon
ENSAE and École Polytechnique
Co-chair, Youth Initiative
2 June 2014

Outline – Measuring Impacts

1. What is **Impact**? Why should we care about measuring impact? How do you really measure impact?

2. How do Randomized Evaluations fit into a broader M&E strategy?

3. Conducting Randomized Evaluations

4. When is an Randomized Evaluation Suitable (or Not)?

5. Additional Resources and Q&A

What is impact and why do we care?

What is impact?

- _Causal_ effect of a program, policy, or a funding decision on an outcome of interest

Why do we care about measuring impact?

- Increases Accountability of the Program
  - _Did the program do what it was supposed to?_

- Improves Development Policy
  - _What is the most effective way to achieve an outcome?_
  - _What are the reasons for success or failure?_
  - _Institutionalize learning and facilitate replications and scale-up_

- Ultimate Goal
  - _Bigger impact on poverty due to more effective programs_
  - _More funding commitment for these proven programs_

A Real Example:
The Problem of Youth Unemployment

- More than 300 million young people worldwide are not in employment, education, or training

- Official unemployment figures understate the problem
    - Do not count underemployment
    - Do not consider quality of employment

How do we improve youth employment?

- Improving their job skills through apprenticeships, internships, training, formal education, soft skills

- Improving matching between youths and jobs through counseling and job search assistance

- Fostering self-employment through entrepreneurship training, finance, access to mentors

- Subsidized employment through labor-intensive public works

But, which of these areas is the most relevant?

And what is the best way to achieve these goals?

Why do we need more rigorous evidence?

- **There is surprisingly little evidence on many important issues in development**

  - Expanding secondary education to all is very expensive. What are the returns to investing in secondary education and what are the relative benefits of technical vs more academic secondary education?

- **Policy makers need to prioritize between different investments**

  - Need to know not just if a policy works but the relative cost effectiveness of alternative investments

But, measuring impact is difficult

- Isolating the impact of a policy or program from all the other changes that are happening when the program is introduced is hard. What changes can we attribute to the program?

- Before/after evaluations compare outcomes before and after the program is introduced
  - Any changes are attributed to the program
  - But many things change over time, not just the program

- Cross sectional evaluations compare outcomes for those who did and did not receive the program
  - Differences in outcomes between participants and non participants attributed to the program

How can we know which of these interventions will work best?

Either look at existing evidence or do a pilot and:

1. Collect anecdotal evidence (people presented to you on visits)

2. Conduct qualitative Surveys (ask local people on surprise visits)

3. Examine before-after difference in wages of participants

4. Simple Comparison: Measure employment rates in school that offered the vocational education program vs. one that did not

5. Difference in Difference of those with program and without

6. Regression Analysis

7. Randomized Evaluations

Measuring Impact –
Anecdotes and Qualitative Surveys

Problems?

1. Anecdotal evidence – cherry pick participants presented to you on visits : there are always success stories, but what about the others?

2. Qualitative Surveys – many biases: surveyed, surveyor or questionnaire – what would have been the situation of participants absent the program
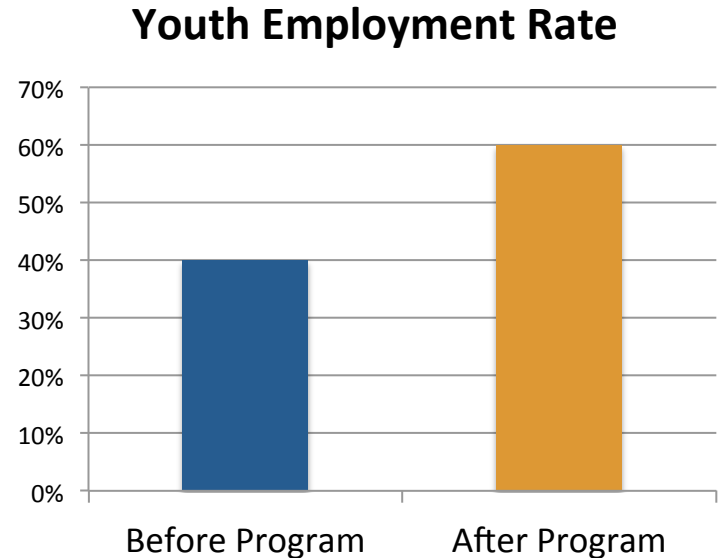
Measuring Impact – *Compare Outcomes "Before" to "After" a Program is Introduced*

What is the Impact here?

– Potential Problems?

How do you disaggregate impact of other things?

- Multinational organization opens new factory

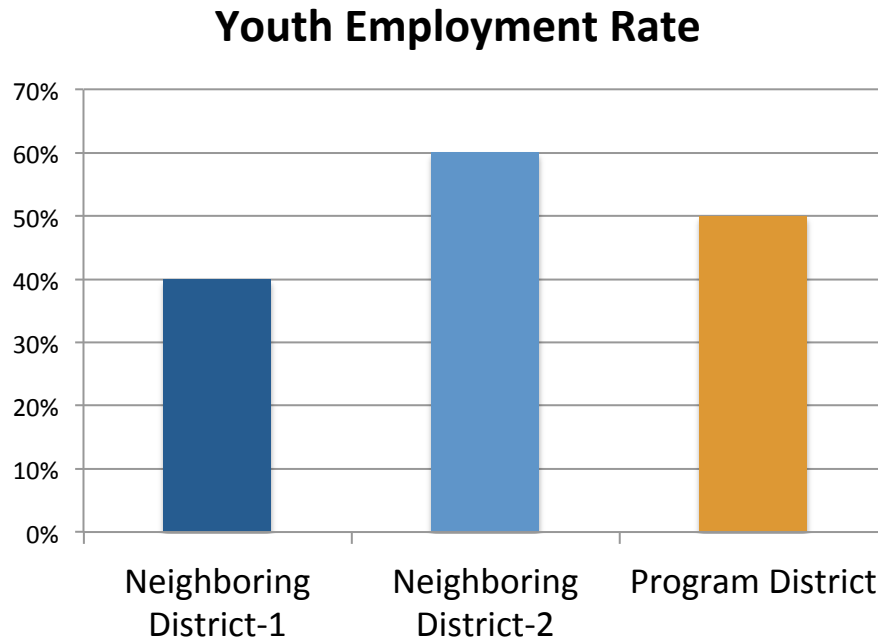- More students graduate or drop out of secondary school

Youth Employment Rate

Measuring Impact – *Simple Difference between outcomes with "Program" vs. "No Program"*

What is the Impact here?

– Potential Problems?

Selection Bias in how district was chosen

- Minister's district with more economic opportunities

- Comparatively well-run schools

- Comparatively poorly run schools

Youth Employment Rate

Regression Analysis

Statistical Tool

– Potential Problems?

Requires Data on Observable and usually Unobservable Variables:

- Quality of district administration

- Motivation of local businesses to hire youth

- Level of community involvement

- Availability of local jobs

- Many characteristics of participants

All these methods attempt to measure impact – *but what exactly is impact?*

Impact is defined as a comparison between:

1. the outcome of participants some time after the program has been introduced

2. the outcome of these participants at that same point in time had the program not been introduced

The *"counterfactual"*
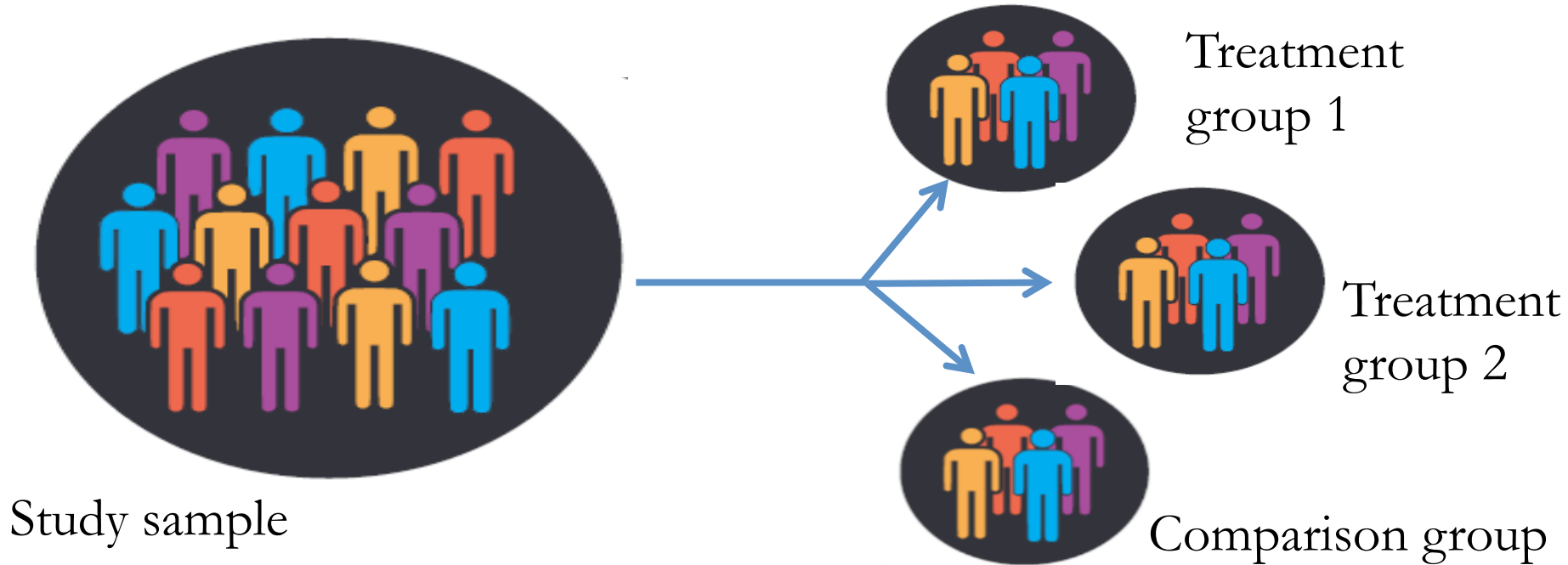
How should we measure impact?

Impact is defined as a comparison between:

1. the outcome of participants some time after the program has been introduced

--and--

2. the outcome of these participants at that same point in time had the program not been introduced (the *"counterfactual"*)

Problem: Counterfactual cannot be observed

Solution: We need to "mimic" or construct the counterfactual

Randomization creates groups with similar characteristics

Study sample
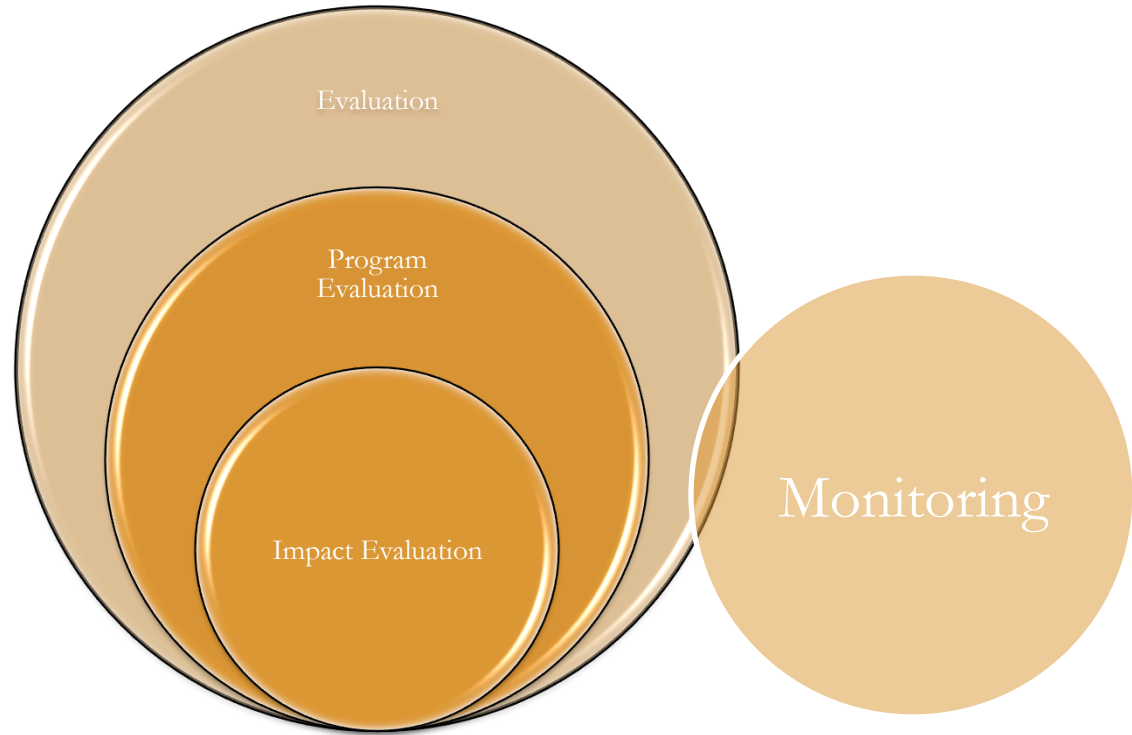
Treatment group 1

Treatment group 2

Comparison group

Outline – Measuring Impact

1. What is **Impact**? Why should we care about measuring impact? How do you really measure impact?

2. How do Randomized Evaluations fit into a broader M&E strategy?

3. Conducting Randomized Evaluations

4. When is an Randomized Evaluation Suitable (or Not)?

5. Additional Resources and Q&A

Monitoring and Evaluation:
What are they and how are they different?

Impact evaluation is one of many components of program evaluation

- Needs Assessment

- Program Theory Assessment

- Process Evaluation

- Impact Evaluation

- Cost Effectiveness

- What is the problem?

- How, in theory, does the program fix the problem?

- Does the program work as planned?

- Were its goals achieved? The magnitude?

- Given magnitude and cost, how does it compare to alternatives?

Needs assessment

Generates a description of the problem and existing solutions using interviews and focus groups, existing data, or new surveys

Example of a question that can be answered:
- What skills are youth lacking that keeps them from accessing jobs?
- Why are vocational programs failing to provide them with the right skills

Program theory assessment

Proposed Intervention: Adapt Vocational Training Program Curriculum to Firms Needs

Unemployment among youth is disproportionately high

> Youth are unemployed because they lack the required skills

> There are sufficient open jobs in their community

> Youth are willing to participate in and complete training program

> Once they have more relevant skills, youth unemployment will be reduced

Process evaluations

Process evaluations involve asking:
- Has the program completed the steps it planned?
- Tracking implementation of each step of theory of change using paper trail and field checks
- Tracking finance

All projects should conduct process evaluations

- No need of a control group: only observe participants

- Qualitative surveys on participants provide objective and continuous feedback

- Consider also surveying implementers in the field

Impact assessment

- **Primary Outcome:**
  - Did vocational training program improve youth skills
  - Did it lead to a reduction in unemployment levels among youth?
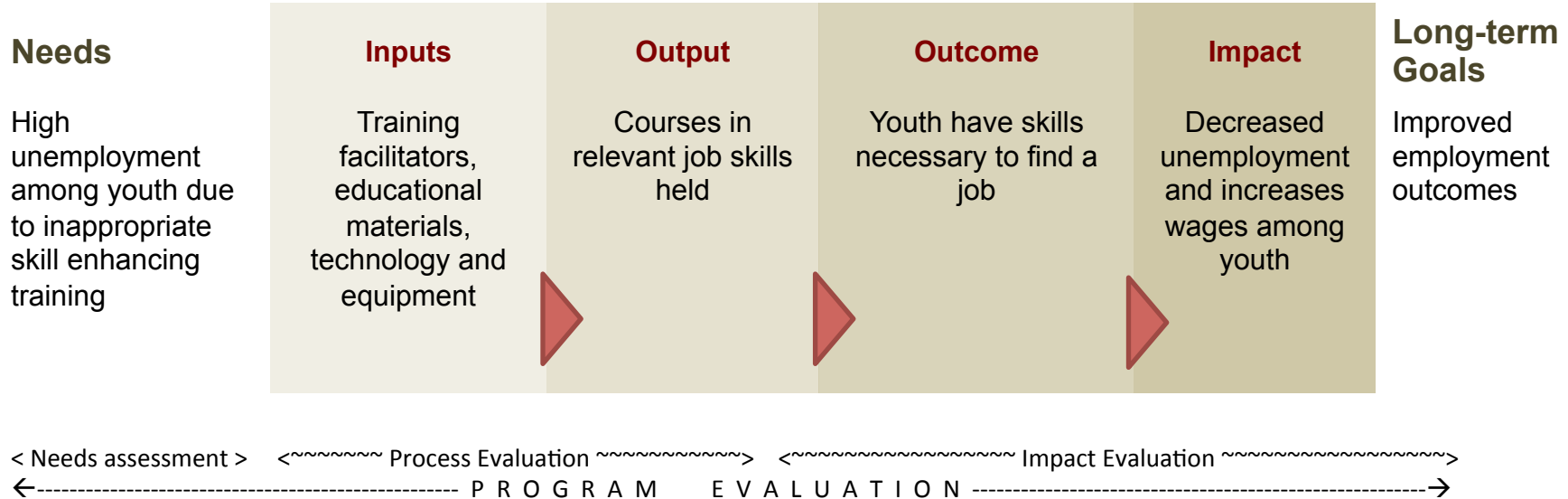 - Did youth earn increased wages after

- **Distributional Questions**:
  - What was the impact for female vs. male participants?
  - How did the training impact highly disadvantaged youth?

- **Long Term Outcomes (if planned):**
  - Did youth stay employed for multiple months or years?

When done well, randomized evaluations involve significant engagement at design, monitoring and evaluation stages

Needs	Inputs	Output	Outcome	Impact	Long-term Goals
High unemployment among youth due to inappropriate skill enhancing training	Training facilitators, educational materials, technology and equipment	Courses in relevant job skills held	Youth have skills necessary to find a job	Decreased unemployment and increases wages among youth	Improved employment outcomes

< Needs assessment >   <~~~~~~~ Process Evaluation ~~~~~~~~~~~>   <~~~~~~~~~~~~~~~~~ Impact Evaluation ~~~~~~~~~~~~~~~~~>

←--- P R O G R A M   E V A L U A T I O N --→
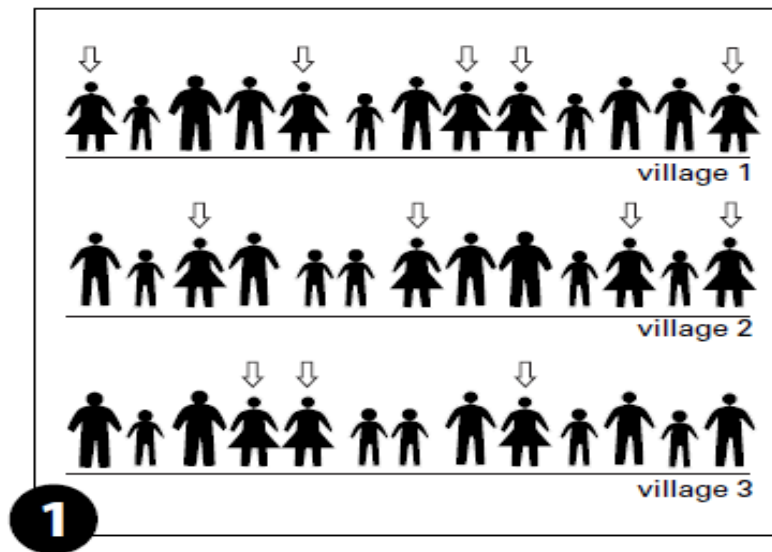
Outline – Measuring Impact

1. What is **Impact**? Why should we care about measuring impact? How do you really measure impact?

2. How do Randomized Evaluations fit into a broader M&E strategy?

3. → Conducting Randomized Evaluations

4. When is an Randomized Evaluation Suitable (or Not)?

5. Additional Resources and Q&A

Random assignment vs. random sampling

- Random assignment
  - People, schools etc. are randomly assigned to different groups (e.g. treatment and comparison)
  - Creates two or more comparable groups
  - Basis of randomized evaluation

- Random sampling
  - Want to measure the characteristics of a group (eg average height)
  - Measure a random sample of the group
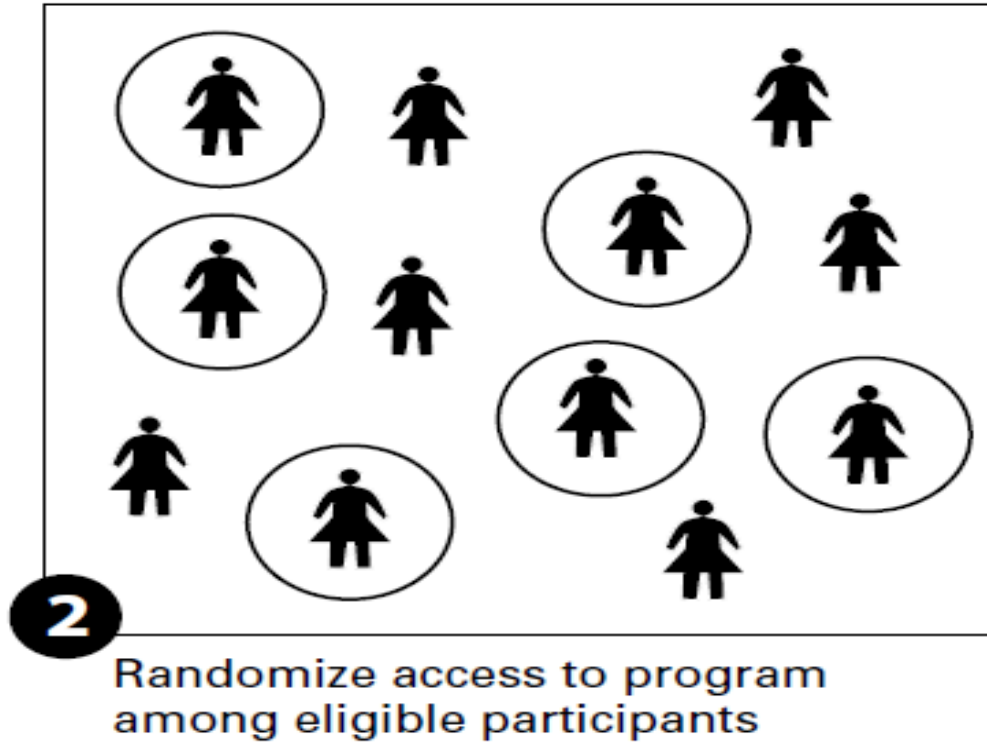 - Often used during randomized evaluations, especially group level randomization
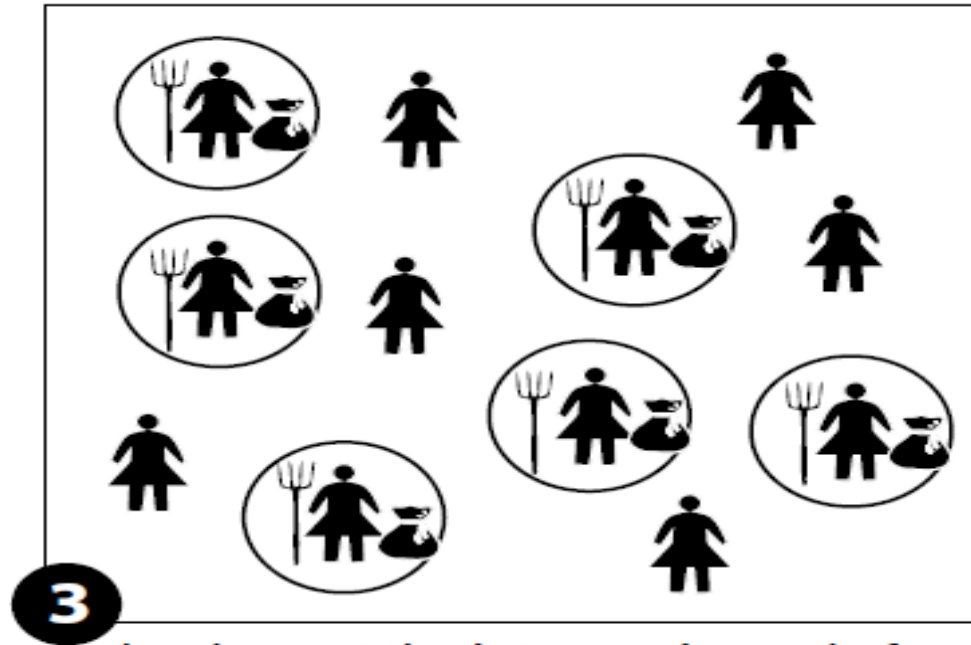
Steps to randomization: 1

1 Define the people who should be eligible for the program

⇒ Micro-credit program available only to young women

Steps to randomization: 2

Randomize access to program
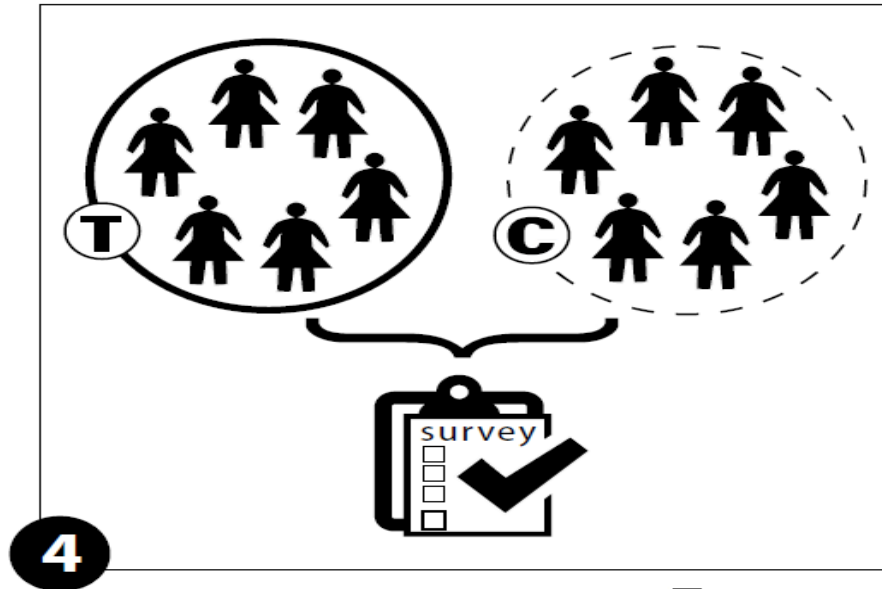among eligible participants

Steps to randomization: 3

3 Implement the intervention only for treatment group individuals

Steps to randomization: 4

Survey people in both the **T**reatment Group (get program) and the **C**omparison Group (no program)

Important design elements of a randomized evaluation

- Sample Size
- Unit of randomization
- Treatment-comparison design variations
- Logistical feasibility

Sample size

- Impact evaluations require a certain sample size to detect a statistically significant effect.

- It is important to think about the evaluations' *statistical power* – its ability to detect a (statistical) effect based on the unit of randomization and treatment-comparison group design and an organization's *implementation capacity*

Unit of randomization

Randomization can occur a different levels for different programs:

- Individual level

- Group level: "Cluster Randomized Trial"
    - School
    - Village
    - County, etc.

How do we choose the level?

- Nature of the Treatment
  - How is the intervention administered?
  - What is the catchment area of each "unit of intervention"
  - How wide is the potential impact?
- Aggregation level of available data
- Power requirements: role of the 'design effect'.
  - power primarily depends on the number of units randomly assigned
- Most natural to randomize at the level at which the treatment is administered.
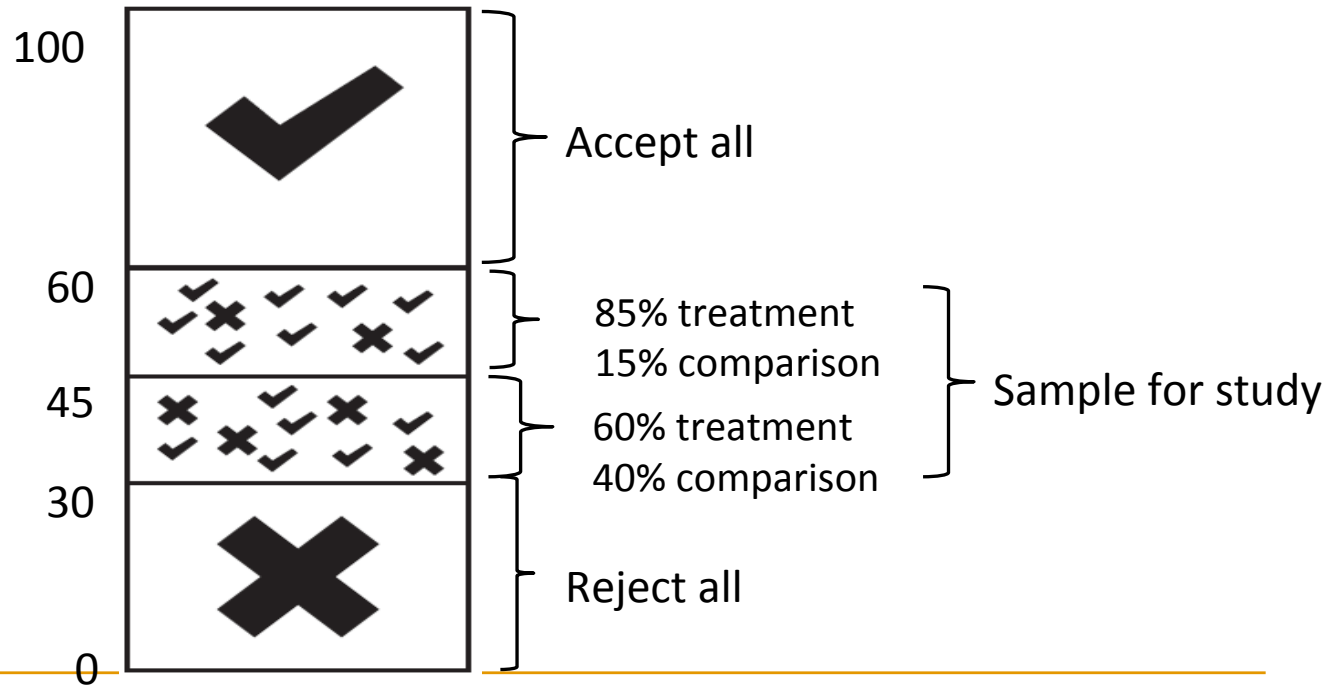
Example: Individual design

- Intervention: Job search counseling for unemployed youth offered through state-run employment assistance centers

- Treatment Level: Individual.

- Randomization level: Individual.

- Unemployed youth from urban sample drawn in 4 cities in Jordan

- Offers of counseling services made directly at the individual level.

Example: Clustered design

- Intervention: Cash transfers for schooling
- Treatment level: Village
- Randomization level: Village
  - Sample of eligible households identified.
  - Households of eligible students in treatment villages receive cash transfer if children remain in school.
  - Power lower than individual treatment, but school monitoring and transfers are both most natural at village level.

Design variations: Randomization around the cutoff

100 — Accept all

60 — 85% treatment / 15% comparison
45 — 60% treatment / 40% comparison
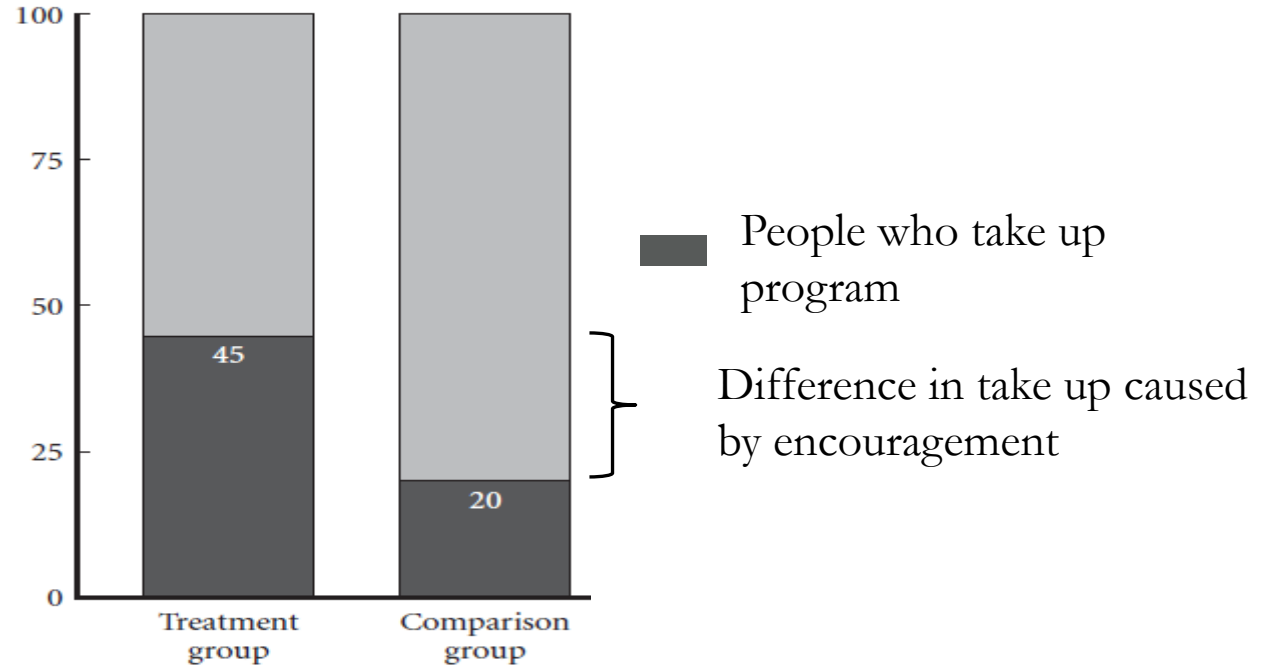30 —

Sample for study

0 — Reject all

Design variations: Randomized phase-in

- Choose target individuals or communities to be covered over several years
- Randomize the order in which they are phased in
- Those not yet phased in are the comparison

Year	Group A		Group B		Group C
Year 1		Treatment group	Comparison group		Comparison group
Year 2		Treatment group		Treatment group	Comparison group
Year 3					

Design variations: Encouragement design

Percentage of people eligible for the program

Treatment group: 45
Comparison group: 20

People who take up program

Difference in take up caused by encouragement

Logistical feasibility

- Is it possible or feasible for staff to implement different programs in the same catchment area?
- Example: Job counselors provides job search assistance and counseling to a number of youth
  - The counselor might serve youth from various treatment and control groups
  - It might be difficult to train them to follow different procedures for different groups, and to keep track of what to give whom

Outline – Measuring Impact

1. What is **Impact**? Why should we care about measuring impact? How do you really measure impact?

2. How do Randomized Evaluations fit into a broader M&E strategy?

3. Conducting Randomized Evaluations

→ 4. When is an Randomized Evaluation suitable (or not)?

5. Additional Resources and Q&A

Ethics of randomization

- When are randomized evaluations ethical?

- Ethics rules:
  - <u>Respect for persons:</u> people are informed of risks and can choose not to participate
  - <u>Justice principle:</u> the type of people who take the risks should also benefit
  - <u>Benefice principle:</u> benefits outweigh the risk of harm (cost) and researchers seek to minimize harm

When to do a randomized evaluation?

- There is an important question you want/need to answer

- Timing: not too early, not too late

- Have time, money and expertise to do it right

When NOT to do a Randomized Evaluation?

- Program is premature and still requires considerable "tinkering" to work well

- Project is on too small a scale to randomize into two "representative groups"

- If a positive impact has been proven using rigorous methodology and resources are sufficient to cover everyone

- After the program has already begun and you are not expanding elsewhere

- If you have weak or no monitoring to ensure that outputs or outcomes are being achieved
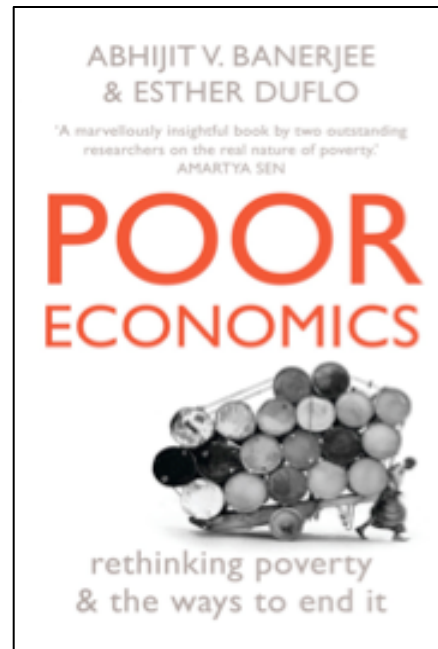
Outline – Measuring Impact

1. What is **Impact**? Why should we care about measuring impact? How do you really measure impact?

2. How do Randomized Evaluations fit into a broader M&E strategy?

3. Conducting Randomized Evaluations

4. When is an Randomized Evaluation Suitable (or Not)?

5. Additional Resources and Q&A

Additional Resources

www.povertyactionlab.org

Evaluations, Policy Lessons, CEAs, Scale-Up Strategies, Publications…

Thank you!

Questions and comments?

Reserve slides

Needs assessment

Generates a description of the problem and existing solutions using interviews and focus groups, existing data, or new surveys

Examples of questions that can be answered:
- What is the level of learning among secondary school students, who is falling behind?
- What skills are youth lacking that keeps them from accessing jobs?

When is a needs assessment useful?

- Focuses attention on a neglected problem

- The problem is sometimes not what we thought
  - Example: in designing a program to provide sanitary pads to help girls stay in school we find
    - they do not skip school more during menstruation
    - Biggest reason for missing school is teacher absence
    - Main complaint about menstruation is period pains

- When we want to design a new program or evaluation
  - Example: before designing an entrepreneurship training program for youth, we need to know how much free time do they have and when? How far can they travel?

Program theory assessment

Proposed Intervention: Vocational Training Program

Unemployment among youth is disproportionately high

> Youth are unemployed because they lack the required skills

> There are sufficient open jobs in their community

> Youth are willing to participate in and complete training program

> Once they have more relevant skills, youth will be able to access jobs and youth unemployment will fall

Process evaluations

Process evaluations involve asking:

- Has the program completed the steps it planned?
- Tracking implementation of each step of theory of change using paper trail and field checks
- Tracking finance

All projects should conduct process evaluations

- Baseline and other surveys provide invaluable information
- Qualitative surveys provide objective and continuous feedback
- Course corrections can be made based on midline surveys
- Implementers can use endline data to change program before scale-up

Impact assessment

- **Primary Outcome:**
  - Did vocational training program reduce unemployment levels among youth?
  - Did youth earn increased wages after

- **Distributional Questions**:
  - What was the impact for female vs. male participants?
  - How did the training impact highly disadvantaged youth?

- **Long Term Outcomes (if planned):**
  - Did youth stay employed for multiple months or years?

Cost-effectiveness or business case assessment

- What does it involve?
  - Examine likely path to impact, what are realistic or best case outcomes. How do those compare to costs?
  - Compare projected cost-effectiveness with alternatives

- Works best when there have been high quality impact evaluations of different alternatives

- When is it sufficient?
  - When even under best case scenario, current project would not be more cost-effective than alternatives
  - Sometimes dividing total cost by number of beneficiaries can be very informative