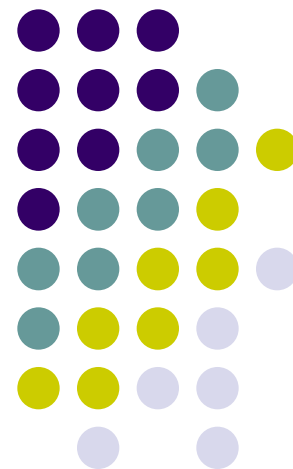
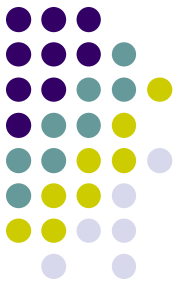


Lessons from Evaluations of Labor Market Programs

Rita K. Almeida
Silvia Paruzzolo
(HDNSP, World Bank)

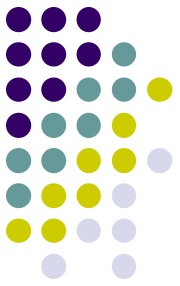




Introduction

- Focus on the evaluation of “active” labor market programs.
- ALMP are designed with different objectives:
 - Improve sector reallocation of workers (retraining programs)
 - Improve skills of those displaced or unemployed (including youth)
 - Insure individuals in the event of job loss (e.g., workfare programs)

Outline Talk



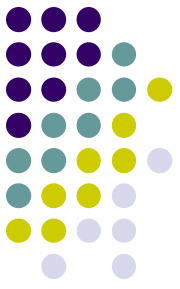
1. Need to Evaluate ALMPs
2. Quantify Program Impact: Role of Impact Evaluations
3. Main Steps of an Impact Evaluation
4. Main Focus: **Impact Evaluation Design**
5. Case studies

Need to Evaluate ALMPs



- Refine program design
- Design new programs
- Improve program targeting (effectiveness)
- Identify inefficient programs (cost-benefit analysis)

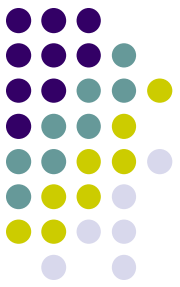
... Usually done through the evaluation of **pilots**.



Evaluation of Pilots

- Impact evaluations are usually done in pilots before programs are implemented/redesigned country wide.
- Pilots focus on smaller sub/groups of the population (e.g., regions, municipalities, groups of the population).

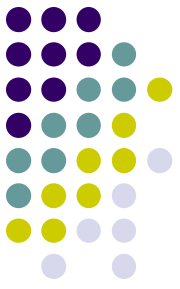
Quantify Program Impact: Role of Impact Evaluations



Monitoring is not Impact Evaluation

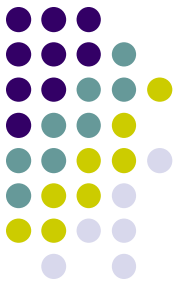
- “Traditional” M&E → NO inherent Causality
 - Is the program being implemented as designed?
 - Could the operations be more efficient?
 - Are the benefits getting to those intended?
- Impact Evaluation → Causality
 - What was the effect of the program on outcomes?
 - Because of the program, are people better off?
 - What would happen if we changed the program?

Main Objective of Program Evaluation

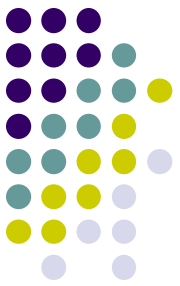


- Estimate the *CAUSAL* effect of program ***X*** on labor market outcome ***Y***
 - E.g., What is the effect of a *job training (X)* on *employability and labor earnings (Y)*?

Main steps for a good IE



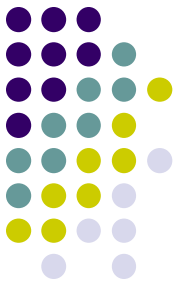
- Policy Relevant Question
- Evaluation Design
- Data collection
- Data Entry
- Analysis.



How to Quantify the Impact?

- Causal effect of program is identified by difference in relevant outcomes of those participating in program and a counterfactual.
- Need to estimate a **Counterfactual**, i.e., what would have happened to participant if he or she had not participated...

Need to Identify the Correct Counterfactual



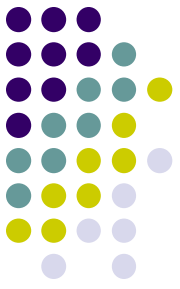
- Could we simply compare:
 - Individuals with and without the program?
 - No. Often individuals could systematically differ in participation and LM outcomes.
 - Same individuals before and after the program?
 - No. Hard to disentangle effect program from other shocks

Problem: Unobserved Counterfactual

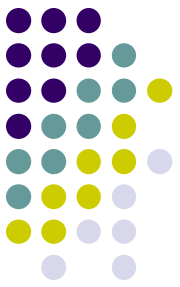


- Treated observation and the counterfactual should have identical factors/characteristics, except for benefiting from the intervention.
 - No other explanations for differences in outcomes between the treated observation and counterfactual.
- **But**, the same person does not exist treated and untreated at the same point in time...

Tools for “Building” Counterfactuals



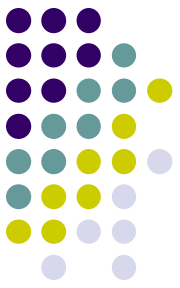
1. Randomized Experiments
- Quasi-experiments:
 2. Regression Discontinuity
 3. Difference-in-difference (Panel data)
 4. Other Methods (e.g., IV, Matching)



Choosing the Methodology

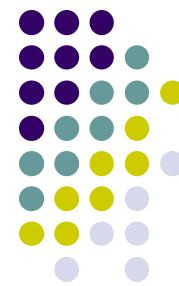
- Choose the most robust strategy (omitted variables, selection) that fits the operational context.
- When possible, explore program budget and capacity constraints:
 - Universe of eligible individuals possibly larger than available resources.
 - Fairest and most transparent way to assign benefit may be to give all an equal chance of participating → randomization

Sampling Basic Requirements



- Must have “large enough” samples (statistical power)
- Must satisfy two requirements:
 - Internal validity (control for all observable sources of difference between treatment and control groups).
 - External validity (expect the same results if you provide the program to different or larger groups).

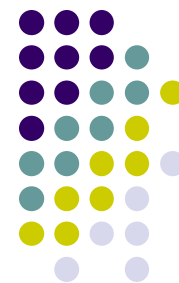
1. Randomization



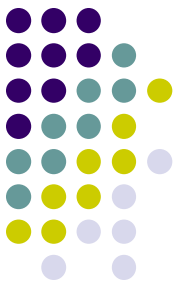
When the eligible population exceeds the number of places:

- **Random assignment (lottery)**
 - Give each eligible unit the same chance of receiving treatment.
 - Compare those offered treatment with those not offered treatment (controls).
- **Randomized phase-in (“pipeline”)**
 - Give each eligible unit the same chance of receiving treatment first, second, third....
 - Compare those offered treatment first with those offered treatment later (controls).

1. Randomization (II)



- But, even randomization can have problems in practice:
 - Political feasibility.
 - External validity/ Scale up challenges: (1) Inputs may change (entry effects);(2) Intervention changes (resources);(3) Outcome (lags; GE).
 - Selective non-participation – Can quantify effect of being illegible but not the effect of program.
 - Need to be careful interpreting results.



2. Difference-in-differences

- Treatment and control groups are systematically different before treatment but reasonable to assume both have **common trends** in the absence of the treatment.
- Compare changes in outcomes over time across treatment group and control group.
- Data Requirement: Need a minimum of three points in time to test assumption (two points before the intervention).

Case Study:

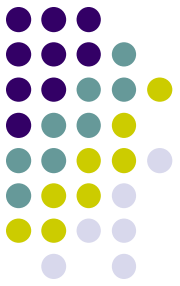
“Jóvenes en Acción” Colombia



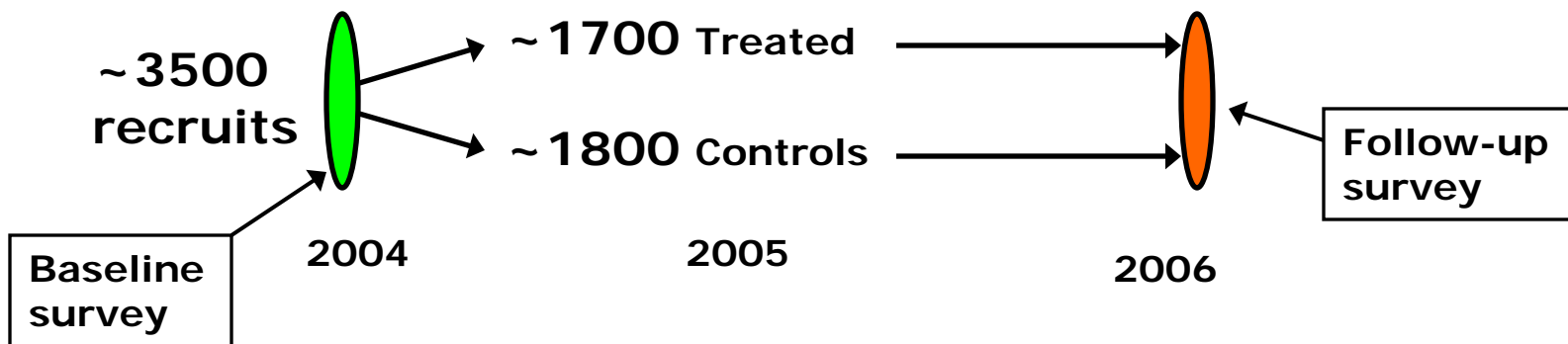
- Beneficiaries: Poor youth (18-25 years old in the 20th lowest deciles of the income distribution).
- Assumption: Main determinant of youth unemployment in country is lack of skills.
- Intervention: Program provided a stipend to trainees throughout the 6 months of program. Training program comprises:
 - 3 months of classroom vocational training – Skills range from cosmetology to the use of computer automated systems.
 - 3 months of on-the-job training- Internship provided by firms in manufacturing, retail and trade, and services.

Source: Attanasio, Kugler and Meghir (2007) Effects of Youth Training in Developing Countries: Evidence from a Randomized Training Program in Colombia; NBER working paper

Case Study: Evaluation Design



- Training providers recruit 50% more candidates than they have room for.
- Participants are randomly selected from recruited candidates.

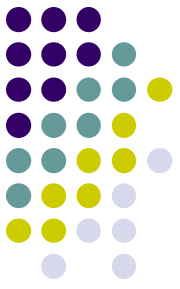


Case Study: Effect Program



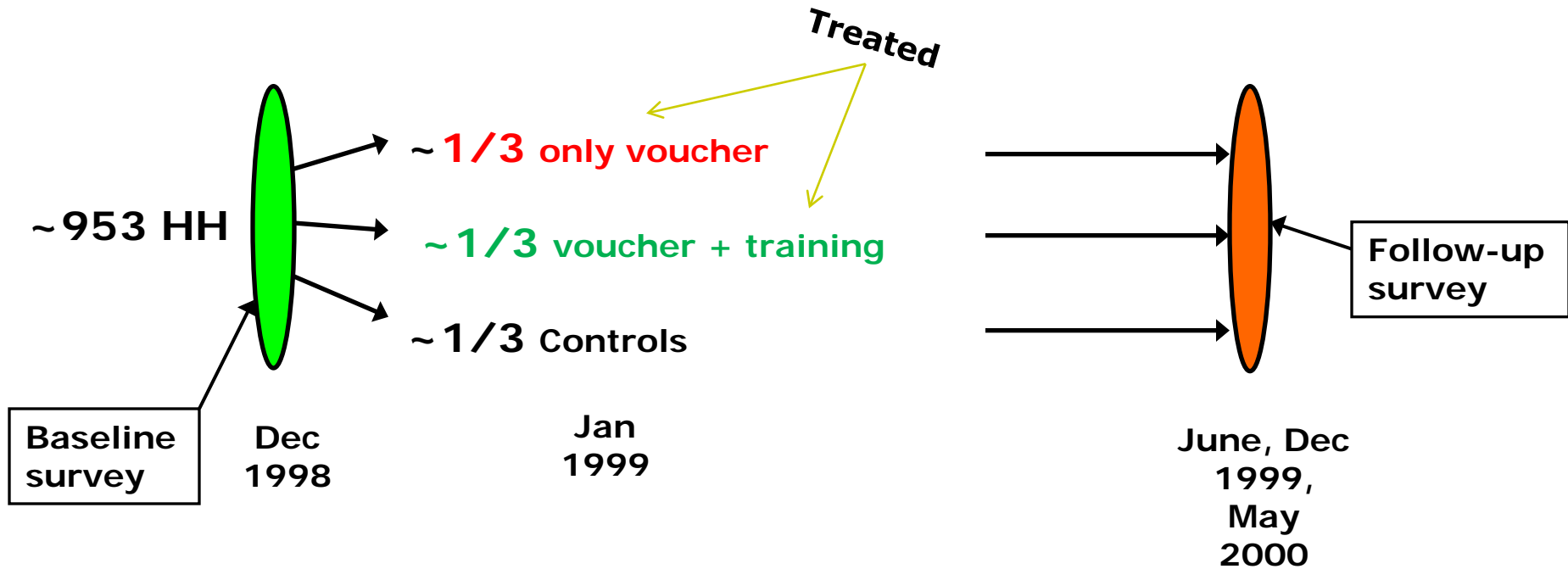
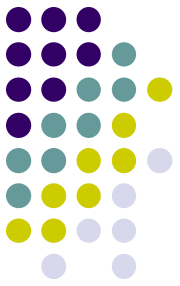
- Training increases labor earnings and employment for men and women. Larger effect on earnings for women than men (18% versus 8%).
- Earnings increases explained by increased employment in formal sector jobs following training.
- Larger effects when individuals spend more time doing on-the-job training, while hours of training in the classroom have no impact.

Case study: “*Pro-empleo*” Argentina



- Beneficiaries: Workfare beneficiaries of *Trabajar* (program granting temporary work, at relatively low wage and oriented to social infrastructure or community services)
- Assumption: workfare beneficiaries do not transition because slack demand or low skills?
- Intervention: *Pro-empleo* provided two instruments to help the transition from welfare to work:
 - Wage Subsidy Voucher
 - Training.

Case Study: Evaluation Design

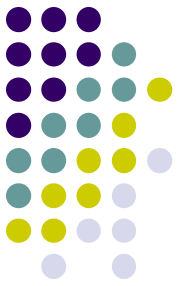


Case Study: Effect Program



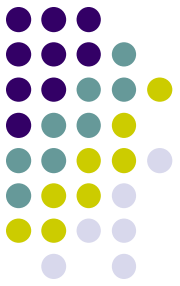
- Wage-subsidy voucher increased the likelihood of wage employment, but did not affect other outcomes (short term).
- Training program had no significant additional effect on any outcome.

Case study: Youth Opportunity Program Uganda



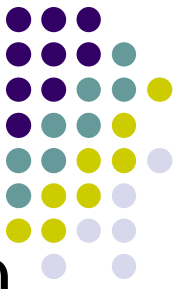
- Beneficiaries: Youth in 29 districts in northern Uganda.
- Question: Is entrepreneurship constrained by lack of skills and/or lack of credit?
- Intervention: Vocational training and Supply of Materials to foster creation of small business

Case study: Youth Opportunity Program Uganda (II)

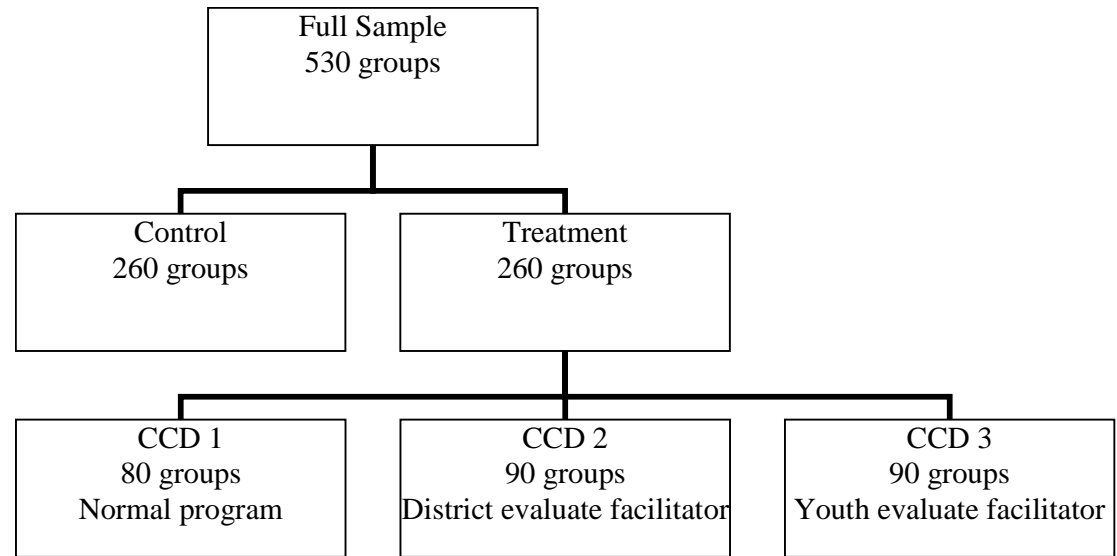


- Anecdotal evidence of poor planning, budget management and little supervision of training provided.
- Complementary test on how to boost the success of training.
- Test the introduction of management, planning, and extension services
 - Management and extension advisor (MEA)
 - \$100 for 6 months of follow-up services
 - Model 1: Groups select MEA; contracting and oversight done by district.
 - Model 2: Groups select MEA; contracting and oversight done by group.

Case Study: Evaluation Design

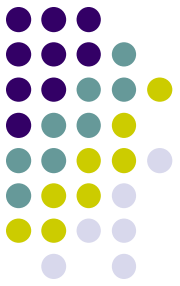


- Random assignment of treatment based on excess demand



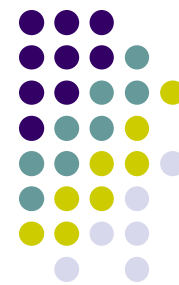
- Outcomes: Labor earnings, employability skills (leadership abilities and empowerment), psychosocial well-being.

Case study: “*Jefes del Hogar*” Argentina



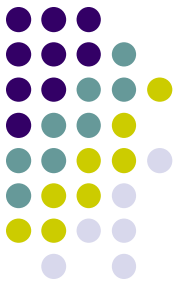
- Beneficiaries: Unemployed household heads with dependents (children aged < 18 or incapacitated) in the aftermath of the Argentinean economic crisis of 2002.
- Question: Did the workfare program help the unemployed reach higher standards of living and avoid poverty?
- Intervention: Cash transfer (150 pesos per month) to participants, which are required to work 20 hours in community work, training activities, school attendance or employment in a private company (wage subsidy). All coordinated at *municipio* level.

Case Study: Evaluation Design

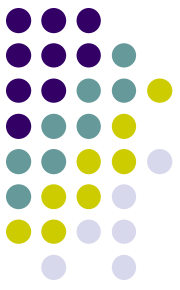


- Comparison group – Program was going through a rapid scaling up. There were many current applicants to the program who had not yet received benefits.
- This group was reasonably similar ex ante. They do not find evidence of selection related to first participants having greater income shocks nor administrative assignment favoring particular groups.
- Use non-experimental methods (Matching) to pair individuals together with longitudinal data (DD).

Case Study: Effect Program

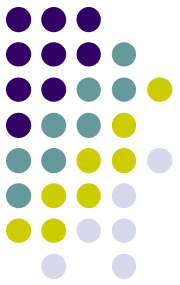


- *Jefes* did reduce aggregate unemployment by attracting individuals to the labor force (both from inactivity and unemployed).
- Evidence suggests that program **did** partially compensate many losers from the crisis and reduced extreme poverty.



Conclusion (I)

- Essential problem of IE- Do **not** observe the outcomes for participants if they had not participated. Evaluation is about missing data.
- A "comparison group" is used to identify the counter-factual of what would have happened without the program.
- Several methods are available (e.g., Randomization, diff-in-diffs, RD, IV).



Conclusion (II)

- Selection should be based on the most robust strategy that fits the operational context.
- Explore program budget and capacity constraints to choose a design.
- But, no method is perfect.... Always desirable to triangulate methods.