

## **Project/program evaluation: different components**

- Monitoring: Did the implementation go as planned?
- Process evaluation: How does the program operate?
- Cost-benefit or cost-effectiveness: How do program costs compare to alternative uses of the same resources and to the benefits produced by the program?
- Impact evaluation: Did the program have the desired effects?  
Are the effects attributable to the program?

## Impact evaluation of (development) intervention

Aim: quantifying the impact of certain intervention on the target population, i.e. “treatment group”

- Quantifying what?
  - Changes in selected indicators of outcomes that can be attributed to a specific intervention
    - e.g. education program: additional years of education, increased likelihood of enrollment, increased likelihood of retention...
- Quantifying how?
  - Identify the specific intervention (program/treatment)
  - Identify/establish the counterfactual (no intervention)
    - Problem: you never observe what would have been the outcome for the beneficiaries if the program had not been implemented
  - Specify indicators of outcomes (~intermediate or final goals)
  - Identify exogenous variables to identify heterogeneity

## Solutions for problem of missing counterfactual

“no-no” solution: ex-post comparison of average outcomes for beneficiaries and non-beneficiaries

problem: selection bias

- Program targeting
- Self-selection into programs

=> Differences in outcomes might be due to other differences between beneficiaries and non-beneficiaries, and not due to program intervention

INSTEAD: Defining appropriate “control” or “comparison” groups

- Experimental design: Randomly allocate intervention among eligible
- Quasi-experimental design: Generate comparison groups that resemble treatment group, at least in observed characteristic, through econometric methodologies

## Experimental Design

Allocate intervention randomly among eligible households/individuals/...

### Advantage

- => Create comparable treatment and control groups that are statistically equivalent to one another (can/should be checked)
- => Quantitative effect of program: easy!
  - = difference between the means of treatment group and control group

### Disadvantage: Potential problems in practice:

- unethical (denying benefits to otherwise eligible hh/ind)
- politically difficult
- scope might prevent it (e.g. if national policy change)
- people moving in and out of control group
- randomness might be difficult to ensure (~e.g. local administrators might have different objectives)

Possible solution: phasing in of program

## Quasi-experimental Design

Generate comparison groups that resemble treatment groups (ex-post)

Advantage: use existing data

- ⇒ quicker and cheaper
- ⇒ evaluate program that was not designed to be evaluated

Disadvantage:

- Reliability (need more assumptions)
- Statistically complex because of selection bias
  - Differences in observables (e.g. characteristics used for targeting)
  - Differences in unobservables (self-selection, political selection,...)

## **Quasi-experiments: Statistical techniques to control for bias**

- Programs with partial coverage

- 1) (Propensity score) matching
- 2) Double difference

- Programs with full coverage

- 3) Instrumental variable
- 4) Reflexive comparison (single difference)

5) Combination of the above

## 1) Matching

Identify non-participants that are comparable in essential characteristics to the participants

- Possible when there exists a large population, which for exogenous reasons has been excluded from the program
- Can be used with single cross-section of data
- Data needed: A sample of participants and a large sample of non-participants from which one can pick the comparison group.
- Construct comparison group by selecting sample that “matches” treatment group
  - Similar set of observed characteristics
  - Similar “propensity score” (predicted probability of program participation given observed characteristics)
    - => propensity score matching
- Calculate difference in outcome between beneficiary and closest match => mean of difference is estimated effect program

## 2) “Double difference” or “Difference in difference”

Compare treatment and control group (first difference), before and after the program (second difference)

- ⇒ Need a baseline survey and a follow-up survey
- ⇒ Can be combined with propensity score matching (to define control group)

Assumption: the difference between before and after in the comparison group is a good counterfactual for the treatment group

- Difference before-after for comparison group represents change in outcome due to trend/other events

$$\bar{y}_{c1} - \bar{y}_{c0} = \frac{1}{N_c} \sum_{j \in C} (y_{j1} - y_{j0})$$

- Difference before-after for treatment group represents change in outcome due to trend/other events and treatment

$$\bar{y}_{t1} - \bar{y}_{t0} = \frac{1}{N_t} \sum_{i \in T} (y_{i1} - y_{i0})$$

⇒ Can isolate the impact of the program:  $(\bar{y}_{t1} - \bar{y}_{t0}) - (\bar{y}_{c1} - \bar{y}_{c0})$

### **3) Instrumental Variables**

Programs with full-coverage but self-selection into program (or placement bias)

Problem: some unobserved characteristics might explain both participation to the program and outcome

- ⇒ Find one or more variables that matter for participation but not for outcomes given participation to isolate the effects (note: this can be difficult to find – sometimes program design can help)

### **4) Simple difference or reflexive comparison**

Compare (trend in) outcomes for population before and after program

- ⇒ Need a baseline survey/pre-program data
- ⇒ Problem: Changes might be due to other factors, not to program

## Further issues related to impact evaluation

- Qualitative methods can be used to inform key impact evaluation questions
  - Details on actual implementation of programs
  - Heterogeneity in outcomes
  - Effects on other than targeted outcome (e.g. forgone income in case of education program)
  - Effects on non-target population
  - Reasons for outcomes
  
- Cost-benefit?
  - Cost of producing the obtained outcome?
  - Learning more from impact evaluation by comparing outcomes of alternative interventions (innovative program designs)
  
- Regression analysis allows for analyzing heterogeneity of outcomes