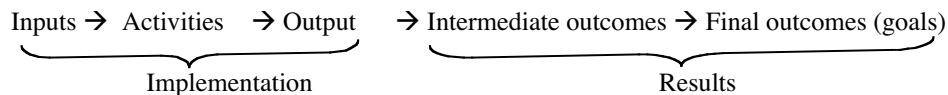


**Handout #6**  
**Impact Evaluation**

**1. Evaluation systems**

***Project sequence:***



***Types of evaluation:***

- *Programmatic evaluation: logframe*  
From activities to outputs and outcomes (indicators).  
Evaluate achieved against planned outputs and outcomes at given times (intermediate and final).
- *Comprehensive expenditure analysis*  
Use of resources: observe and explain inconsistencies between actual and planned expenditures.
- *Impact analysis:*  
Changes in selected indicators of outcomes that can be attributed to a specific intervention.

To do an impact analysis:

- We need to clearly identify a specific intervention (what program, what expected objectives, at what time, at what place, applied to what unit of analysis).
- We need to identify a counterfactual with no intervention against which the change with intervention can be measured: before/after, with/without.
- We need to specify indicators of outcomes (endogenous variables) to be used to measure impact. Hence, the project objectives (goals, mandates) need to be clearly defined. These indicators must be observable before/after or with/without the intervention. They can be indicators of intermediate or final outcomes.
- If there is impact heterogeneity, we need to identify exogenous variables that may make the impact differentiated across units of analysis.
- We need reliable/credible/verifiable information.

***Objectives of evaluation systems:***

Often required by law: Yearly in Mexico, as required by Congress; U.S. 1993 Government Performance and Results Act, fully implemented starting in 1997.

Allows to engage in results-based management. Use results of evaluation to:

- Assess value of program (ex-post).
- Adjust program (feedbacks): minor adjustments, major adjustments, redesign, cancel.
- Link to resource allocation, budgeting, personnel management.
- Evaluation is a learning process (hence role of participation, ownership).
- Improving evaluation = learning to learn (start simple, use pilots, and improve over time).
- Need incentives to learn, use results, and change programs.

**Impact evaluation challenge and techniques:**

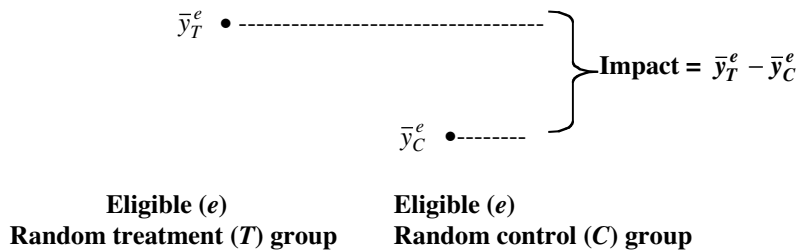
- Selection bias:
  - program placement
  - self-selection
- Techniques for impact evaluation
  - Experimental design, randomization. Treatment and control groups
  - Quasi-experimental design: Treatment and comparison groups
    - Matching methods, Double-difference techniques
  - Non-experimental design,
    - Instrumental variables
    - Statistical methods
  - Qualitative methods

**2. Experimental design – Randomization**

Randomization allows to create identical treatment and control groups.

- Procedure and ethical issue
  - Treatment group and control group
  - Example: Rural education program Progresa in Mexico

- Program impact from simple difference



$$\text{Impact} = \underbrace{\frac{1}{N_T} \sum_{i \in T} y_i}_{\text{average outcome in treatment group}} - \underbrace{\frac{1}{N_C} \sum_{j \in C} y_j}_{\text{average outcome in control group}}$$

Can be done on subgroups to evaluate heterogeneity of program effect

Example: School subsidy in urban Pakistan (Quetta)

Need to check that control and treatment groups have similar distribution of exogenous variables, outcome prior to program (if available), and behavior prior to the program (if available).

### 3. Matching method to construct comparison groups

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

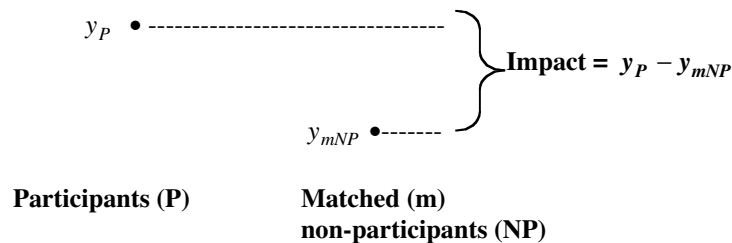
Possible for program with partial coverage, i.e., when there exists a large population which, for exogenous reasons, has been excluded from the program.

Key: The selection of the participants is not related to the outcome of the program

Examples: Local programs

By contrast to:

- credit program placed where economic opportunities are highest.
- health clinics placed where most needed
- self-selection for program participation



- Data needed: A sample of participants (usually from a special survey designed for the program evaluation) and a large sample of non-participants (usually some other large existing survey, such as LSMS for households) from which one can pick the comparison group. Both surveys must include variables  $X$  that are important determinants of program participation and outcome.
- Simple matching: Construct a comparison group with non-participants that have characteristics that are “similar” to those of the participants.  
Example: Job training program for women, unemployed for 3 less than 3 months, and without kids.
- Propensity score matching (individual matching):  
Variables  $X$  that help predicts program participation.

Instead of matching on all the  $X$ , one matches on the probability of participation.

- a) Using both sample to estimate the probability of participation as function of variables  $X$ .
- b) For each participant  $i$  to the program find the closest matches  $m(i)$  among the non-participants, i.e., the non-participants with the closest predicted value of probability of participation. One can choose 1 to 5 closest matches.
- c) For each participant  $i$ , compute the average outcome of the closest matches:

$$y_{mi} = \frac{1}{n} \sum_{j \in m(i)} y_j$$

- d) The impact of the program is:

$$\text{Impact} = \frac{1}{N_T} \sum_{i \in T} (y_i - y_{mi})$$

There are many variations of this method (using just one match, or a few matches, or many matches that are weighted according to how “close” they are to the treated person).

Example: Argentina’s workfare program Trabajar

Note: The income effect is less than the payment from Trabajar, because of the foregone income. By subtracting the predicted income effect from each observed income, one can estimate the “without program” income.

Whatever the matching method, program impact can be computed for different subgroups of the population to assess heterogeneity in the impact of the program.

#### 4. Double difference method

When the control or comparison groups are not perfectly comparable:

- imperfect randomization, imperfect matching
- there is no group with comparable characteristics

- *Double-difference over time:*

Data needed: Observations before and after the implementation of the program, for both the treatment group and the comparison group.

Hence a base line survey and a follow-up (evaluation) survey

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

- a) Compute the difference before-after for the comparison group:

$$\bar{y}_{C1} - \bar{y}_{C0} = \frac{1}{N_C} \sum_{j \in C} (y_{j1} - y_{j0}).$$

Represents the change in outcome due to natural trend and all other events.

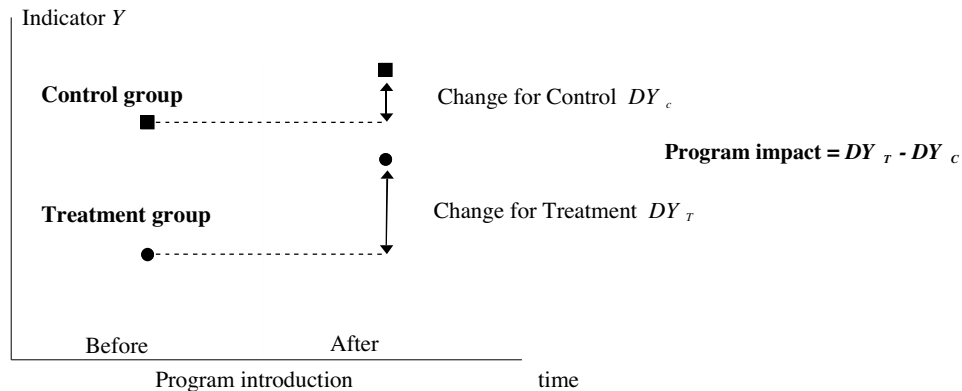
- b) Compute the difference before-after for the treatment group:

$$\bar{y}_{T1} - \bar{y}_{T0} = \frac{1}{N_T} \sum_{i \in T} (y_{i1} - y_{i0}).$$

Represents the change in outcome due to natural trend and all other events, and the program

c) The impact of the program:

$$\text{Impact} = (\bar{y}_{T1} - \bar{y}_{T0}) - (\bar{y}_{C1} - \bar{y}_{C0})$$



Example: School subsidy program in urban Pakistan (Quetta) (imperfect randomization)

- *Double-difference with an excluded group (in cross-section data)*

Data needed: Eligibility criteria for participation to the program. Observation of eligible groups in treatment and control areas. Observation of ineligible groups in treatment and control areas.

Example: Grameen bank: eligible households with less than 0.5ha, non-eligible households with more than 0.5ha. Both groups observed in villages with Grameen Bank, and in villages without Grameen Bank

Example: Progresa: eligible poor households and non-eligible non-poor households in villages receiving Progresa and in villages not receiving Progresa.

Assumption: the difference between non-eligible and eligible households in the non-treated areas is a good counterfactual for the same difference in the treatment areas.

a) Compute the difference between the eligible and non-eligible groups in the control areas:

$$\bar{y}_{Ce} - \bar{y}_{Cn} = \frac{1}{N_{Ce}} \sum_{j \in Ce} y_j - \frac{1}{N_{Cn}} \sum_{j \in Cn} y_j$$

average outcome  
among eligible
average outcome  
among eligible

Represents the “natural” difference in outcome between eligible and non-eligible.

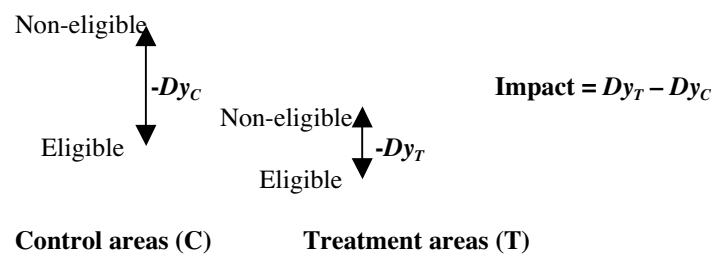
b) Compute the difference between the eligible and non-eligible groups in the treatment areas:

$$\bar{y}_{Te} - \bar{y}_{Tn}$$

Represents the difference in outcome between eligible and non-eligible, when the eligible groups benefit from the program.

c) The impact of the program:

$$\text{Impact} = (\bar{y}_{Te} - \bar{y}_{Tn}) - (\bar{y}_{Ce} - \bar{y}_{Cn})$$



Examples: Progresa

These impact values are called “Average Treatment Effect”

- *Heterogeneity:* To take into heterogeneity, the double difference can be done by regression technique:

Without heterogeneity, the results previously obtained will also be found by regressing the outcome on dummies of group:

For each observation unit  $i$ , note:

$\delta_i = 0$  if observation  $i$  is from the base line,  $\delta_i = 1$  if it is from follow-up.

$T_i = 1$  if in the treatment group, and  $T_i = 0$  if in the comparison group.

Then regress outcome  $y$  on  $\delta$ ,  $T$ , and the product  $\delta T$ :

$$y_i = a + b\delta_i + cT_i + d\delta_iT_i + \varepsilon_i$$

Check that:

$$\bar{y}_{T1} = a + b + c + d$$

$$\bar{y}_{T0} = a + c$$

$$\bar{y}_{C1} = a + b$$

$$\bar{y}_{C0} = a$$

therefore  $\text{Impact} = d$ .

To take into account the heterogeneity in characteristics, add control variables  $X$ :

$$y_i = a + eX_i + b\delta_i + cT_i + d\delta_i T_i + fX_i\delta_i T_i + \varepsilon_i$$

Example: Progresa

To account for heterogeneity of impact in the population, add control variables  $X$  in interaction with the impact:

$$y_i = a + eX_i + b\delta_i + cT_i + d\delta_i T_i + fX_i\delta_i T_i + \varepsilon_i$$

The impact of the program is then:  $\boxed{\text{Impact} = d + fX_i}$  which varies with values of  $X$ .

## 5. Controlling for unobservables without comparison groups: Instrumental variables

When the program has universal coverage (every one is offered the program) but not all individual choose to participate (self-selection bias). The same technique can be applied when the program has been offered in areas with particular needs/characteristics (placement bias).

Example: El Salvador's EDUCO Program of community-managed schools. EDUCO school requires parents' direct participation.

In the regression of the achievement of the students:

$$y_i = a + bX_i + cT_i + \varepsilon_i$$

the treatment variable (being in an EDUCO school) is endogenous.

- Bias due to endogeneity (self-selection or placement bias):  
When some unobserved characteristics explain both participation to the program and outcome.  
For example only parents more conscious of the importance of education choose to send their children to EDUCO schools. Their children would do better in either type of schools. What is captured in the parameter  $c$  is not only the effect of the program but also the effect of "parents' education consciousness" on the performance of the children.
- Suppose participation to the program can be explained by exogenous variables  $Z$ :

$$T_i = d + eZ_i + \mu_i$$

These two equations can be jointly estimated by the Instrumental Variable method. When the participation is estimated with a probit model (because it is a variable that take the value 0 and 1 only), a similar method is employed called the Heckman

method. These two methods correct for the fact that there may be some unobserved variables, such as ability, vitality, “parents’ education consciousness”, etc., that both explain the choice to participate in the program and the value of the output. Since they are not observables they are part of the “errors”  $\varepsilon_i$  and  $\mu_i$ , which are therefore correlated.

- The IV methods requires that at least one variable in  $Z$  explains the choice to participate to the program, and does not affect the outcome, i.e., does not pertain to  $X$ . This is not always easy to identify.

### 6. Evaluation of full-coverage interventions: reflexive comparison and statistical controls

Example: Impact of Nafta on the exports to Mexico. We have observations for several years before and observations for several years after the 1994 implementation of Nafta.

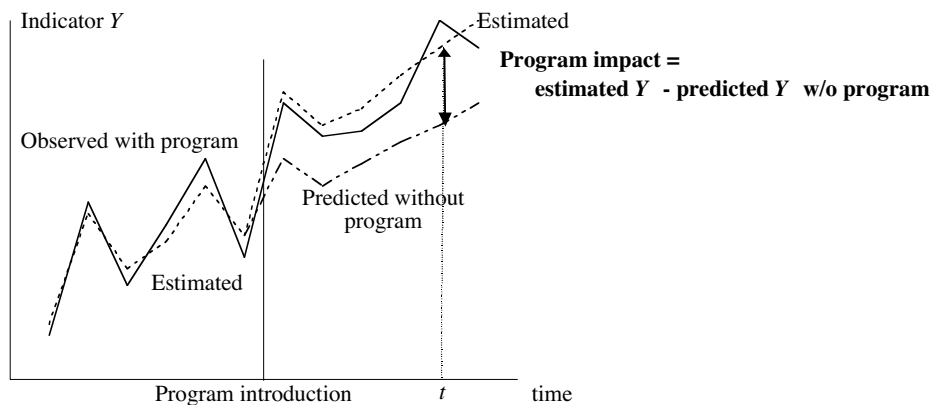
- Reflexive comparison consists in comparing exports before Nafta and exports after Nafta. The problem is that it does not separate the Nafta effect from the devaluation and the crisis that took place in Mexico at the same time.

- Statistical control:  
This is done by the estimation of:

$$E_t = a + bX_t + cT_t + \varepsilon_t$$

where  $T$  represent Nafta (= 0 before 1994 and 1 after 1994) and  $X$  include control variables that explain exports, such as GDP in Mexico and in the US, real exchange rate between Mexico and the US, etc.

Therefore **Impact = c**. It measures the difference in predicted outcome with and without the program.



## 7. Qualitative methods

To gain insight from beneficiaries and program administrators on what works and does not work.

To identify what are the reasons for the choice to participate in the program.

Complementarity with quantitative methods:

- Done before the analysis, and even preferably before the survey is conducted:  
Used to help form hypotheses for the quantitative methods.  
Used to help define the information that need to be collected  
Used to identify the main dimensions of heterogeneity of impact
- After the quantitative analysis or preferably in interaction:  
Used to assess the plausibility of the results, and their interpretation  
Used in the process of refining hypotheses and analysis

## 8. The special problem of impact assessment for CDD programs

### 1. Descriptive analysis

Impact analysis of CDD poses the special problem of choice of objectives by the agent:

A strong municipality may choose a difficult objective that it may fail to achieve.

A weak municipality may choose easy objectives that it will likely fulfill.

Objectives chosen by the municipality (agent) may differ from those that the principal desires.

This does not create a problem for descriptive impact analysis. Impacts are measured:

- According to the objectives (indicators) chosen by the principal.
- According to the objectives (indicators) chosen by the agent if they are known.
- Using a set of indicators of agent behavior selected by the analyst if the agent's objectives are not known.

### 2. Positive impact analysis

Endogenous choice of objective creates a problem for the causal analysis of impact. We need to:

- First explain the choice of objective (for which we need a valid instrument).
- Second explain performance for the chosen objective.

### 3. Normative impact analysis

Endogenous choice of objectives creates a problem for normative impact analysis (placing a judgment about the quality of impact).

We can distinguish the following two situations:

i) The objectives of the agent are the same as those of the principal; or the objectives of the agent are irrelevant and the program is to be assessed in terms of the objectives of the principal (he who pays).

Appraise impact according to this criterion.

ii) The objectives of the agent are different from those of the principal

Alternatives are:

- Appraise impact according to the principal's objectives (as above).
- Appraise impact according to the agents' objectives if they are known. However, if different agents have different objectives, their performances cannot be compared.
- If the agents' objectives are not known, can only proceed by imposing one's own criteria. However, this implies judging agents' behavior according to criteria that are not theirs.

#### References for examples:

\*Kim, Jooseop, Harold Alderman, and Peter F. Ozarem. 1999. "Can Private School Subsidies Increase Enrollment for the Poor? The Quetta Urban Fellowship Program." *The World Bank Economic Review*, Vol. 13, No. 3, pp. 443-65. (Random placement, although ex-post treatment and control groups are not identical; Double difference; selection on observables).

Jalan, Jyotsna, and Martin Ravallion. "Estimating the benefit incidence of an antipoverty program by propensity-score matching. *Journal of Business & Economic Statistics*. 21(1):19-30, 2003 Jan (Propensity Score Matching).[http://web2.infotrac-custom.com/pdfserve/get\\_item/1/S6c6815w2\\_1/SB533\\_01.pdf](http://web2.infotrac-custom.com/pdfserve/get_item/1/S6c6815w2_1/SB533_01.pdf)

Jimenez, Emmanuel, and Yasuyuki Sawada. 1999. "Do Community-Managed School Work? An Evaluation of El Salvador's EDUCO Program." *The World Bank Economic Review*, Vol. 13, No. 3, pp. 415-441. (Heckman)

Lyons, Karen. "NAFTA after 5 years: A Study on the Increase of Bilateral Trade between Mexico and the United States. Undergraduate Honor Thesis, UC Berkeley, Dpt of Economics, 1999.

Raymond, Melanie. "Schooling and Migration in Rural Mexico" PhD dissertation, 2001, UC Berkeley, Dpt of Ag. and Resource Economics.