PP 253 / ARE 253 Sadoulet / de Janvry
Fall 2007

Handout #6
Impact Evaluation

1. Evaluation systems

Project sequence:

$$\underbrace{\text{Inputs} \rightarrow \text{Activities} \rightarrow \text{Output}}_{\text{Implementation}} \rightarrow \underbrace{\text{Intermediate outcomes} \rightarrow \text{Final outcomes (goals)}}_{\text{Results}}$$

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 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.
- We need to identify a counterfactual with no intervention against which the change with intervention can be measured: before/after, with/without.
- Need to have data for many units of observations to do statistical analysis.

***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

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

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 that, for exogenous reasons, has been excluded from the program.

Key assumption for the validity of the method: Selection on observables: Once you control for all the observable characteristics, the participation to the program is not correlated to any other determinant of the outcome.

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.

- 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:

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

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.

- The matching method is similar in spirit to a simple regression model in which you "control for" the observable characteristics with a pre-defined function: Estimating

$$y_i = a + eX_i + \delta T_i + \varepsilon_i$$

identifies the impact of the treatment *T* if ε is orthogonal to *T*, i.e., conditional on the observables *X*, participation *T* is not correlated to the other factors influencing the outcome. This regression is however more restrictive than the matching method as it imposes a given function *y* (*X*).

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

Selection on unobservables: the participation decision is based on some unobservable characteristics that are correlated to the outcome of interest. In the regression framework this would be equivalent to say that ε is correlated with T.

4.1. *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

Key assumption for the validity of the method: 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:

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

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

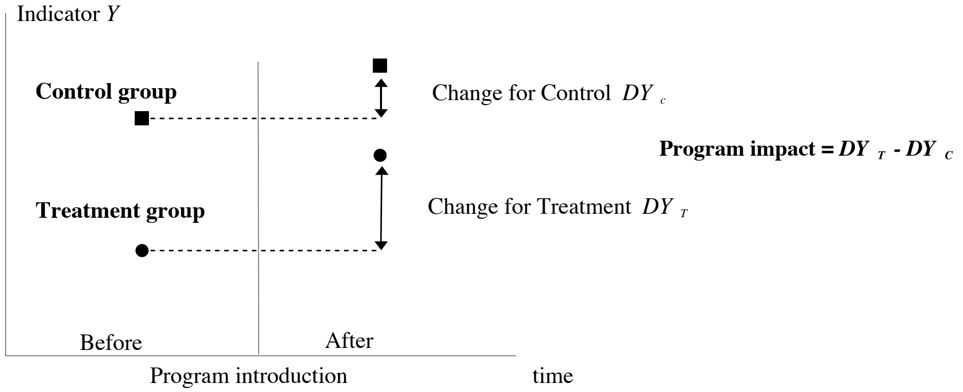
b) Compute the difference before-after for the treatment group:

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

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

c) The impact of the program:

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



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

The result will also be found by regression analysis as follows:
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 δ, T, and the product δT:

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

Check that:

$$\bar{y}_{T1} = a + b + c + d$$
$$\bar{y}_{T0} = a \quad\quad + c$$
$$\bar{y}_{C1} = a + b$$
$$\bar{y}_{C0} = a$$

therefore $\boxed{\text{Impact } = d}$.

4.2. 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.

Key assumption for the validity of the method: 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:

$$\bar{y}_{Ce} - \bar{y}_{Cn} = \underbrace{\frac{1}{N_{Ce}} \sum_{j \in Ce} y_j}_{\text{average outcome among eligible}} - \underbrace{\frac{1}{N_{Cn}} \sum_{j \in Cn} y_j}_{\text{average outcome among eligible}}$$

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

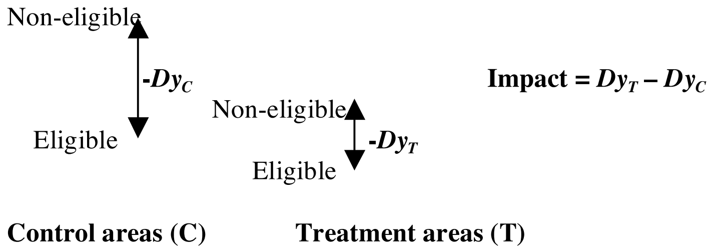
b) Compute the difference between the eligible and non-eligible groups in the treatment areas:

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

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

c) The impact of the program:

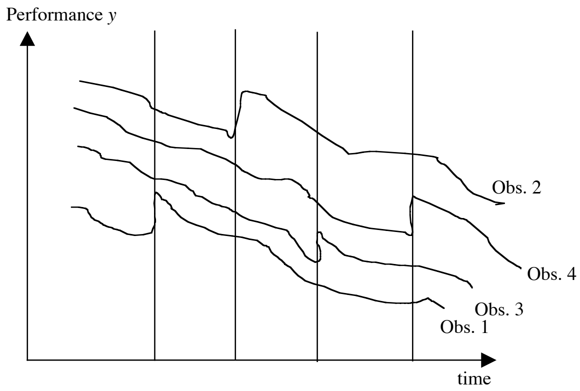
$$\boxed{\text{Impact } = \left(\bar{y}_{Te} - \bar{y}_{Tn}\right) - \left(\bar{y}_{Ce} - \bar{y}_{Cn}\right)}$$



Examples: Grameen Bank (Morduch), School construction in Indonesia (Duflo)

4.3. Generalization of the double-difference: Staggered entry in panel data

When difference units enter the treatment at different points in time.



Key assumption for the validity of the method: The entry into the treatment is not correlated with the performance of the unit.
(Counter-example: Each year, you choose the schools with the worse performance to implement a remedial program)

The analysis is done in a regression framework with panel data:

$$y_{it} = \delta T_{it} + \mu_i + \nu_t + \varepsilon_{it}$$

where the unit fixed effect μ_i account for the cross-sectional difference in performance (the level), the time fixed effect ν_t for whatever changes over time is common to all units (on the graph, the downward trend), and T_{it} is the treatment variable, equal to 1 after unit *i* has entered the program, and 0 before.

Example: Privatization of water supply in Argentina (Galiani, Gertler, Schargrodski)

4.4. Heterogeneity of impact

The previously calculated impact values are called "<u>Average Treatment Effect</u>"
To account for heterogeneity of impact in the population, add control variables *X* in interaction with the impact in the regression framework:

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

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

5. Controling 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).

<u>Example</u>: 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" ε_i and μ_i, which are therefore correlated.

Key sssumption for the validity of the method: there exist some exogenous factor Z that influences the participation to the program but not directly the outcome

- 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. 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 needs 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

References for examples:

Duflo, Esther. “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy experiment.” *American Economic Review*, Sept 2001, 91:4, 795–813. (Double difference, varying exposure to program, instrumental variables)

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodski. 2005. Water for Life: The Impact of the Privatization of Water Services on Child Mortality. *Journal of Political Economy*. 113(1). [Staggered entry in panel data].

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].link

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)

*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).

Morduch, Jonathan. "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh". mimeo, June 1998. (double difference).