Impact Evaluation Methods

The impact analysis approach

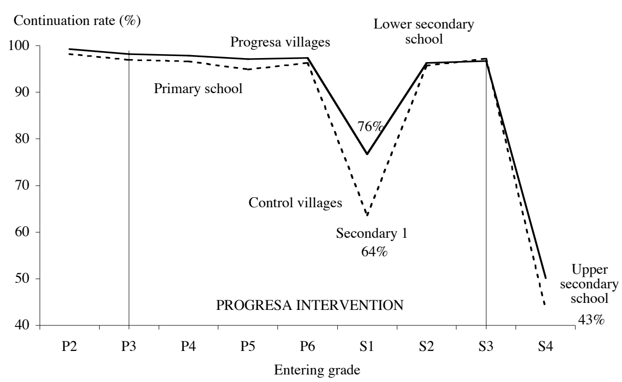
- Focus on one variable (the "treatment", e.g. voucher, change in tariff, training, etc.) and ignore the rest
- Design the sample in a way that would make the direct comparison of individual that benefit from the program and those that do not sufficient to infer the **causal effect of the treatment**.
- The ideal would be to … randomize the program over a large sample of individuals. Whenever possible and meaningful, just do it.
- If not possible, not meaningful, then … find a **counterfactual** situation that allows you to answer the question:

  How would individuals who benefited from the program have fared in the absence of program?

  Those situations are called "counterfactual".
- The focus is on the design of the strategy to establish a counterfactual.
- Econometric technique usually simple

1. Experimental design – Randomization

de Janvry, Alain, and Elisabeth Sadoulet. 2006. "Making Conditional Cash Transfers More Efficient: Designing for Maximum Effect of the Conditionality." *World Bank Economic Review* 20(1): 1-29.
Progresa: Conditional cash transfer in Mexico



Key assumption: If it were not for the treatment, the control and the treatment population would be *statistically* identical, i.e., have identical expected values for the outcome of interest, whatever the treatment value *T*: $E\left(y_i \mid i \text{ in Treatment group}, T\right) = E\left(y_j \mid j \text{ in Control group}, T\right)$

Program impact from simple difference:

$$\text{Impact} = \underbrace{\bar{y}_T}_{\substack{\text{average outcome}\\\text{in treatment group}}} - \underbrace{\bar{y}_C}_{\substack{\text{average outcome}\\\text{in control group}}}$$

To obtain standard errors, done with the regression:

$$y_i = a + \delta T_i + u_i$$

with the key SLR4 assumption: $E\left(u_i \mid T_i = 0\right) = E\left(u_i \mid T_i = 1\right) = 0$ that ensures that $\hat{\delta}$ is an unbiased estimator of δ

This is based on a subsample so results are not identical to the graph:

```
. reg enroll98 program if grade98==6;

      Source |       SS       df       MS              Number of obs =    2172
-------------+------------------------------           F(  1,  2170) =   47.28
       Model |  9.43680066     1  9.43680066           Prob > F      =  0.0000
    Residual |  433.153899  2170  .199610092           R-squared     =  0.0213
-------------+------------------------------           Adj R-squared =  0.0209
       Total |    442.5907  2171  .203864901           Root MSE      =  .44678

------------------------------------------------------------------------------
    enroll98 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     program |    .135874   .0197613     6.88   0.000      .097121     .174627
       _cons |   .6306197   .0155737    40.49   0.000     .6000788    .6611606
------------------------------------------------------------------------------
```

Test for the validity of the randomization

We cannot test $E\left(u_i \middle| T_i = 0\right) = E\left(u_i \middle| T_i = 1\right) = 0$, but we can test for the equality of means of the observable characteristics, with the idea that if there is no statistical difference in the observable characteristics, it is plausible that there is no difference in the unobservable as well.

Test that for any variable not affected by the program: $E\left(x_i \middle| i \text{ in Treatment group}\right) = E\left(x_i \middle| i \text{ in Control group}\right)$ cannot be rejected (and is small, meaning that large differences are rejected).

	Control	Treatment	p-value for test of equality between T and C
Male	0.49	0.51	0.428
Household head age	44.5	43.9	0.179
Household head is male	0.93	0.93	0.495
Indigenous	0.42	0.39	0.125
Household head education	2.68	2.64	0.701
Household size	7.45	7.50	0.615
Number of rooms	1.78	1.71	0.109
Has a bathroom	0.63	0.60	0.128
Distance to school (in kms)	2.35	2.42	0.437

Intention-to-Treat Effect: Estimating the impact of risk on farmer behavior – Using RCT

Emerick, Kyle, Alain de Janvry, Elisabeth Sadoulet, and Manzoor Dar. "Technological innovations, downside risk, and the modernization of agriculture "

In 2011, we selected 128 villages prone to flood in Orissa, India. We randomly divided them in 64 "treatment" villages and 64 "control" villages. In each of the villages, we randomly selected 5 farmers. In the treatment villages, we randomly selected 5 of the farmers and gave them a small package of a new variety of rice seeds that are flood resistant. We let them cultivate as usual, save their seeds as usual and reuse them if they wanted the next year as usual, but also give/trade some with friends if they wanted. 2011 was a flood year and hence a good opportunity to learn about the value of this new seed. At the end of 2012, we survey the 10 farmers in each village and obtained the information on their rice cultivation.

Note that some of the "treated" farmers did not reuse the new variety in 2012 (2 farmers did not use the seeds we gave them in 2011 and 77 did not reuse them in 2012, out of the 320 initial recipients), some of the "control" farmers obtained new variety seeds from others (21 in the control villages and 42 in the treatment villages). But this is endogenous, so we keep the initial allocation of treatment. The estimator is called "Intention to Treat".

Verification of the balance of characteristics

Table 2: Mean values of predetermined characteristics by treatment status

	Original control	Original minkit recipient	P-value of difference
Land owned in hectares	0.81	0.868	0.18
HH has private tubewell	0.332	0.325	0.82
HH has piped water	0.035	0.057	0.14
HH has refrigerator	0.078	0.076	0.92
HH has television	0.628	0.605	0.5
Education of farmer	6.896	6.946	0.83
Age of farmer	51.191	51.783	0.46
HH has thatched roof	0.557	0.548	0.79
HH has latrine	0.289	0.354	0.08
HH has electricity	0.843	0.822	0.42
HH has below poverty line card	0.574	0.559	0.67
Schedule Tribe / Schedule Caste	0.189	0.176	0.59

Data are from year 1 follow-up. Values in columns 1 and 2 are means. P-values in column 3 are based on t-tests of equality of means. Standard errors are adjusted for clustering at the village level.

Impact of reduced risk on farmer's behavior

Farmer's behavioral response to reduced risk

		Plot level behavior			
	Fertlizer (kg/ha)	Use Swarna	Use traditional variety	Broadcast instead of replanting	Yield (kg/ha)
Original minikit recipient	24.6**	-0.102***	-0.041**	-0.063***	283.5***
	(10.00)	(0.017)	(0.016)	(0.017)	(77.48)
Mean of Dep Variable (kg)	215.49	0.36	0.28	0.19	2818
Number of Observations	1235	4578	4577	4571	4573
R squared	0.615	0.116	0.27	0.243	0.16

Source: Various tables in Emerick et al. (2013)

All regressions include: Scheduled tribes/castes, Hh has below poverty line card, HH has thatched roof, and block fixed effects. Standard errors are adjusted for clustering at the village level.

Adding covariates and heterogeneity

We can use observable covariates Z to add precision to the estimation and to verify, as a robustness check, that $\hat{\delta}$ is invariant to the introduction of covariates in the regression

$$y_i = \alpha + \delta T_i + Z_i \beta + \varepsilon_i .$$

We can also measure heterogeneity of the program effect for individuals with specific characteristics B (such as gender, age, socio-economic status, etc.) by interacting these characteristics with the treatment variable. We now estimate:

$$y_i = \alpha + \delta T_i + Z_i \beta + \gamma B_i + \phi T_i B_i .$$

In this equation, the impact on an individual i with characteristics B_i is $\hat{\delta} + \hat{\phi} B_i$.