

# **Some Reflections on the Practice of Program Evaluation in LAC**

Sebastian Galiani  
Washington University in St.  
Louis

## **The Domain of Evaluation**

- It may be self evident that programs should regularly be assessed and revised in the light of the lessons drawn from experience.
- Nevertheless, systematic program evaluation is a recent development.
- But what to evaluate, and how?

## **The Domain of Evaluation**

- To be useful in policy formation, evaluations should answer *counterfactual* questions.
- Most evaluations focus only on impact questions. Also, evaluations invariably consider only a small set of program options. This is an important area for further research.
- For example, in the case of conditional cash transfers, it has been claimed that making the transfers to mothers (instead of HH) was one of the causes of the success of these interventions. This was not evaluated though.

## **The Domain of Evaluation**

- Other example, also related to CCTs, is the conditionalities themselves. What are they causal effects? (being currently evaluated in Morocco).
- Selection into programs and take up rates could also be subject to counterfactual examination. This is an area of research that has been practically ignored in LAC. Addressing this issue could also help us to generalize the results of specific instances of program interventions.
- A good example is the paper by Heckman and Smith (2004) on the determinants of participation on the JTPA.

## **Heckman and Smith (2004)**

- This article presents a framework that decomposes participation in the Job Training Partnership Act (US) into multiple stages:
- Eligibility, awareness, application and acceptance and enrollment.
- This decomposition is useful in order to understand the sources of inequality in the receipt of social programs.

## **Decomposing the Process of Selection**

- In order to determine at what stage and in which direction particular observable characteristics operate to determine participation in the program, the authors use the chain rule to decompose the probability of participation.
- They find strong evidence that informational barriers play a major role in determining differences in program participation rates.

## **Evaluation Methods**

- Program evaluation have the pragmatic objective of learning what works best.
- However, the reasons program works as they do are rooted in basic social processes: Individual labor supply and fertility behavior, the operation of the labor market, the formation of social norms; the way states, localities, and service providers organize and administer the programs, etc.
- What must one understand about how programs work in order to reach conclusions about their merits?

## **Reduced-form and Structural Evaluation**

- Both approaches have their uses. In general, reduced-form evaluation is feasible if programs have actually been operated in the environment of interest.
- Structural evaluation is feasible if one is able to characterize the environment of interest and if one understands the social process at work well enough to permit forecasting with confidence.

## Reduced-form and Structural Evaluation

- Structural evaluation demands stronger assumptions than reduced-form experimental evaluation.
- Needless to say, in all cases, we economists want to learn why and how things work.

## Pragmatism in Program Evaluation

- Nowadays, a common critique to reduced-form evaluations is that it does not (**always**) recover structural parameters.
- Suppose that  $Y = F(x, w)$ , and that we can experimentally manipulate  $x$ , but that  $x$  also affects  $w$ .
- A reduced-form evaluation will only estimate the total derivative of  $Y$  with respect to  $x$ .
- Econometrically, via a structural model, we could attempt to separate the direct and indirect effects (the one operating through  $w$ ) of  $x$  on  $Y$ .
- However, even if these two effects were identified, to conduct cost/benefit analysis we will also need to know an ancillary model, namely, that of  $W = g(x)$ .

## **Pragmatism in Program Evaluation**

- Yet, another instance in which we might want to rely on a reduced-form model for policy evaluation is one where the structural model is of difficult use in the design of interventions.
- Consider the case of cognitive achievement of Children. Say there is production function  $Y = F(x_1, x_2, \dots, x_n)$ .
- Even if identified, we might not know how to affect  $x_1, \dots, x_n$ . Of course, this opens a research agenda, but, also, it might justify the use of reduced-form causal parameters such as:
- $\Delta = E[Y(I=1) - Y(I=0)]$

## **Pragmatism in Program Evaluation**

- We might be considering two alternative interventions though: I1 and I2. This poses a much difficult question because when administered over the same individuals, they might have a different average effect than when administered independently.
- This appears to be the case of early child interventions over the life cycle. In this situation, relying on structural evaluation might be preferred (see the recent work of Heckman and Todd and Wolpin).
- However, still, in light of the discussion in our previous slide, we would like our model to accommodate these interventions (or variables that we can be mapped from them).
- Providing experimental answers to these types of questions requires a multifactorial randomized design.
- Other important case where the explicit use of a structural model might be preferred is when there are equilibrium effects

## **Better data**

- I got to be convinced that one of the limitations in LAC to advance more the structural evaluation agenda is the lack of suitable datasets.
- There is nothing nearly comparable to the National Longitudinal Surveys of Labor Market Experience or the PSID, which in the U.S., are then merged with administrative datasets.
- Administrative data is substantially underused.

## **Generalizing Causal Effects**

- There is a sense in which all causal generalization is about interpolation and extrapolation.
- This is clearly easy to do with an structural model than with reduced-form models. Nevertheless, it also would most likely entail the need for stronger assumptions.
- In the latter case, Rubin (1992) propose estimating response surface (coarsely stated, it implies plotting the function relating a cause to an effect across a broad range on some third variable. I find this idea very useful for generalizing the effects of program evaluations.

## Quasi-experiments

- Finally, I do not believe that reduced-form evaluation needs only to rely on experimental designs. Randomized evaluations are preferred (even with the caveats that has been extensively discussed in the literature).
- However, I am still sympathetic to quasi-experimentation. Here, the question is to explain plausible rival hypothesis. If possible, one should not restrict oneself to a single research design or approach when trying to assess the impact of a program.
- Any and all methods that help to control or explain plausible rival hypothesis are useful.

## Quasi-experiments

- When constructing a causal hypothesis one should envisage as many different consequences of its truth as possible (falsification tests (e.g., Duflo (2001), Angrist and Kruger (1992) and Berlinski, Galiani and Gertler (2006)).
- Also, rely on Multimethods when it is possible –Lavy (2002), Berlinski and Galiani (2007) and Berlinski, Galiani, McEwan and Shapiro (2007).