



# Social Safety Nets Primer Notes

2003 ■ No. 14

## Impact Evaluation of Social Programs: A Policy Perspective

Governments and donor organizations increasingly recognize that rigorous evaluations of public interventions should be part of the social policy decision-making process. Yet there is frequently a gap between the desire for information on the effectiveness of programs and an understanding of the potential and the limitations of evaluation tools. This note reviews the basic elements of good impact evaluations, identifies some of the political economy aspects that influence whether they are conducted, and explores ways to encourage use of evaluation.

### Elements of An Impact Evaluation

Program-impact evaluations assess the impacts on participants attributed to direct participation in a program or intervention. It should not be confused with other types of policy assessments or with program monitoring. Good evaluations include not just quantitative estimates of program impacts, but also seek to explain why they occurred (or did not), and what the resulting policy implications may be. Full program evaluations have evolved to include several typical components: a) a *process study*, which examines the particular program operations and processes; b) an *impact study*, the technical heart of the evaluation; and c) a *cost-benefit assessment* that compares the costs of offering the program with the benefits to participants.

Conducting a successful impact evaluation requires advance planning. There are five major considerations.

**Selecting a specialized evaluator** – external to the government or implementing agency – is generally preferred to ensure objectivity and independence. In addition, quantitative impact evaluation requires specialized skills and expertise.

**Selecting quantitative methods** to estimate impacts fall into two strategies. Experimental estimates compare the outcomes of the participants with those of a randomly assigned control group that is otherwise eligible for the program and similar to the participants, but did not receive program benefits. Experimental design is

preferred on methodological grounds because it minimizes the effects of pre-existing differences between the participants and the comparison group that can be confounded with the effects of program participation (selection bias).

If random assignment is ruled out, however, it may still be possible to estimate impacts reliably using non-experimental methods. Multivariate-regression models, matched-comparison methods, double-difference and instrumental-variables methods can attempt to control statistically for sources of selection bias. Most recent impact evaluations in developing countries have relied upon non-experimental methods due to cost and data-availability considerations.

**Integrating qualitative methods** can add rich detail and permit a grounded analysis of the underlying causes of outcomes. They permit a deeper understanding of program processes, external conditions, and individual behaviors. The methods are open-ended, relying on semi-structured interviews in an individual or group setting and interviewer observations.

### Key Design Features of a Good Impact Evaluation

To provide the highest value, an impact evaluation should include:

**Clear objectives.** Evaluation questions should be determined early, be simple and measurable.

**Credible evaluator.** The evaluator should be independent of the agency or institution whose program is being evaluated.

**Rigorous methodology.** Experimental estimates are the ideal but a well-chosen matched comparison group may suffice.

**Adequate sample size.** The sample should be large enough to detect program effects of plausible size. In addition, the size should permit assessment of program impacts on key subgroups of the target population, as appropriate to the program. Minimum detectable effects should be determined prior to the implementation of the evaluation.

**Baseline data.** Need to establish the appropriate comparison group and to control for observable program selection criteria.

**Sufficient follow up.** Follow-up data should be collected after enough time has passed to plausibly detect an impact, and should measure the relevant outcome variables.

**Multiple evaluation components.** The impact evaluation should do more than detect program effects – it should also examine program process, reasons for observed outcomes, and cost effectiveness.

*Source:* Blomquist (2003), Ezemanari et al. (1999).

**Data availability and quality** are the most important factors affecting the quality of analysis. Often, new surveys are required to get sufficient information on program participants, including baseline and follow-up surveys. However, existing household data can sometimes be used for non-experimental methods. For example, more than 30 countries have a version of the Living Standards Measurement Survey (LSMS), and many have surveys for multiple years.

There is considerable variation in the cost of impact evaluations of social safety net and other social programs, ranging anywhere from US\$ 200,000 to over US\$ 1,000,000, with an average for a rigorous evaluation probably between US\$ 300,000 – 400,000. Features affecting total costs include the number and type of policy questions to be addressed, the methodology, extent of new-data collection, the size and scope of the program being evaluated, and local capacity.

### Political Economy of Evaluation

Politics and political economy play an important role in the decision to conduct a program evaluation. The issues stem from principal-agent problems, where stakeholders – including variously the government or funding agency, the implementing agency, and beneficiaries – do not have consistent incentives to support an evaluation.

One reason for reluctance comes from a misunderstanding of what an evaluation can deliver. Specifically, results are not always available in time for short-term decisions and they can appear ambiguous and difficult to translate into policy actions. The fact that different evaluations of the same program can produce conflicting results depending on data sources and methodology seems at first glance to limit the policy value of the exercise. For example, two separate evaluations of Peru's Social Fund, completed a year apart and using different methodologies and data, arrived at opposite conclusions on key impacts. However, this and other possible examples only reinforce the point that evaluation design is critical. The best evaluations use an experimental design and/or use a variety of methods to estimate impacts, thereby providing an indication of the robustness of the findings.

A second factor hindering the use of formal evaluations is political concern over their conduct and the possible repercussions from the results. Evaluation is assumed to be very costly, particu-

larly in relation to the scarce resources available for social programs. Negative findings have the potential to hinder social agendas and damage political careers. It may seem easier and safer for policymakers not to present their detractors with a club to beat them.

However, political need can also work in favor of evaluation. One of the main reasons formal impact evaluations are undertaken is to gain political support for a program. This is particularly true for programs that are seen as strategically important for national policy, or for programs that are introducing innovative approaches. The Mexican government paid for the evaluation of Progresá in part because the conditional cash-transfer model was relatively new and viewed as a potential replacement for certain subsidy programs.

Evaluation can provide improved general knowledge about a type of program, yet, as with public goods, that often is not an adequate incentive for the individual policymaker or government to undertake a specific evaluation. Outside donors and agencies might consider providing more resources for country-program evaluations or create incentives for program evaluation and assessment.

### Developing an Evaluation Culture

International experience suggests that more could be done to foster an evaluation culture. There are several possibilities.

'Policy entrepreneurs' might assist in the education and outreach, **convincing stakeholders** by the demonstration effect of actually doing an evaluation. Funding agencies can make evaluation mandatory. Some governments have followed this strategy, imposing the evaluation requirement from a level of authority that cannot be challenged by the program proponents, as in Chile and the U.S.

Including stakeholders in the evaluation process and disseminating findings is very important to **ensuring transparency and communication**. This will help gain support and ensure results are actually used for policy.

Many of the countries where evaluation has taken hold are those that have put a priority on **developing institutional capacity**, as in Chile during the early 1990s. Donors can make support for evaluations part of the underlying project to help build local capability.

The World Bank Social Safety Nets Primer series is intended to provide a practical resource for those engaged in the design and implementation of safety net programs around the world. Readers will find information on good practices for a variety of types of interventions, country contexts, themes and target groups, as well as current thinking on the role of social safety nets in the broader development agenda.



World Bank, Human Development Network  
Social Protection, Social Safety Nets  
<http://www.worldbank.org/safetynets>

WORLD BANK INSTITUTE  
Promoting knowledge and learning for a better world



Printed on 100% post-consumer recycled paper