

*Impact Evaluation of Social Funds***An Introduction***Laura B. Rawlings and Norbert R. Schady*

I. WHY CARRY OUT IMPACT EVALUATIONS?

Despite the importance of knowing whether development programs achieve their objectives, impact evaluations remain rare in developing economies. This is unfortunate. With the growing use of results-based management by governments, determining whether goals have been attained and convincingly linking changes to specific programs has become increasingly critical. Tracking such outcomes as gains in school enrollment or reductions in infant mortality is indispensable. But simply gathering good data on outcomes sheds little light on why objectives have or have not been met. For this reason, impact evaluations should be a key instrument in policymakers' monitoring and evaluation toolbox.

Impact evaluations rely on the construction of a counterfactual—an attempt to estimate what a given outcome would have been for the beneficiaries of a program if the program had not been implemented. Impact evaluations thus address causality and allow results to be attributed to specific interventions. The challenge of evaluation research arises from the fact that the counterfactual outcome is inherently unobservable, because people cannot simultaneously participate and not participate in a program. The four social fund evaluation studies in this issue illustrate that establishing a counterfactual is usually a matter of using statistical or econometric techniques to construct a control or comparison group.

II. WHY EVALUATE SOCIAL FUNDS?

Social funds have become popular vehicles for channeling development assistance, with a reputation for implementing community-based development projects quickly and with broad participation. That reputation led to their rapid expansion after the creation of the first social fund in Bolivia in 1987.

By May 2001 the World Bank had financed more than 98 social fund projects in 58 countries. Almost all countries in Latin America and the Caribbean have social funds or development projects that embody many of their operational

Laura B. Rawlings is Senior Monitoring and Evaluation Specialist and Norbert R. Schady is Senior Economist, both in the World Bank's Latin America and the Caribbean Region. Their e-mail addresses are lrawlings@worldbank.org and nschady@worldbank.org, respectively.

© 2002 The International Bank for Reconstruction and Development / THE WORLD BANK

characteristics. Many countries in Africa, Asia, Europe, and the Middle East have also established social funds.¹ These social funds have absorbed more than US\$8 billion in investments by the World Bank, other international agencies, and governments.² Nonetheless, at the national level social fund spending remains relatively limited, with total expenditures usually accounting for no more than 1 percent of gross domestic product (Rawlings and others 2002).

The scope and financial scale of the global portfolio of social funds, along with the renewed interest in community-based development models, sparked demand for the evaluation of social funds among national governments and development institutions alike.

III. WHAT CHALLENGES DO SOCIAL FUNDS POSE FOR IMPACT EVALUATION?

Social funds pose particular challenges for impact evaluation. Perhaps the two most important ones relate to the nonrandom placement of projects. Most social funds use a poverty map and other tools to target their investments. But allocation rules may rely at least in part on characteristics of communities that are observed by the social fund administrator but not by the evaluator. If these characteristics affect outcomes, the evaluator's inability to adequately control for them can bias the results. For example, if a social fund spends more in poor communities and poverty is correlated with poor health status and imperfectly observed by the evaluator, estimates of the social fund's impact on health outcomes may be biased downward.

The second challenge arises because communities generally self-select into social fund projects. This, too, can bias the results. For example, if a social fund spends more on school infrastructure in communities that have greater (unobserved) organizational capacity and that in the absence of the project would have been more likely to find other solutions for decaying school infrastructure, estimates of the social fund's impact on educational attainment may be biased upward.

Also complicating the impact evaluation of social funds is the range of objectives they address. This makes selecting valid outcome indicators difficult and comparing outcomes across countries even more so. Despite these challenges, the four evaluation studies in this issue shed light on the ability of social funds to reach poor communities and households and the impact of social fund investments on a number of outcomes.

1. For periodically updated information on social funds, go to the World Bank Social Fund database Web site at <http://worldbank.org/sp> (click on the link for "social funds"). Questions about the social fund Web site or database can be directed to the Social Protection Advisory Service, 1818 H Street NW, Room G8-138, Washington, D.C. 20433 (phone 202-458-5267; fax 202-614-0471; e-mail socialprotection@worldbank.org).

2. Of the \$8 billion in investments in social funds, the World Bank accounts for about \$3.5 billion. This total excludes social funds that do not receive World Bank financing and are instead financed by other multilateral and bilateral sources or solely through domestic resources.

IV. WHAT METHODS DO THE FOUR EVALUATIONS USE?

The articles that follow present four different approaches to the same basic challenge: determining whether social fund investments have led to changes in the well-being of beneficiaries.³ The evaluations have used a range of methods, including randomization, propensity score matching, and instrumental variables. Some have relied on several approaches, which can provide a useful check on the robustness of the assumptions underlying different estimates. In each case the choice of evaluation methods reflects available data, time, and resources as well as the particular focus of the different evaluations.

Randomization

Randomization assigns the “treatment”—in this case a social fund intervention—through some sort of lottery, allowing researchers to construct treatment and control groups. In a study of the Bolivian social fund, Newman and others use randomization of the offer to participate in a social fund project to evaluate the impact of improvements in school infrastructure on a variety of school outcomes in the rural Chaco region.

Randomization is immensely appealing because if the sample is large enough, this method controls for all differences, observable and unobservable, between the treatment and control groups. Simple differences in mean outcomes between the two groups, or differences in changes in outcomes, can then be credibly interpreted as the impact of the treatment on the treated.

But the evaluation of the Bolivian social fund also shows some of the limitations of randomization and some of the challenges that social-sector programs pose for evaluation. The evaluation had to deal with changes that occurred after the allocation of the offer to participate, as a result of which some schools not selected for treatment ended up receiving the social fund intervention. Newman and others use bounds estimates to correct for these changes.

Techniques Matching

Propensity score matching in its simplest form involves predicting the probability of treatment on the basis of observed covariates for both the treatment and the control group samples. This probability, the propensity score, is then used to match treated and untreated observations—for example, through nearest-neighbor matching. Under some conditions the difference in mean outcomes between the two groups is then a reasonable estimate of impact.

Propensity score matching is often fairly simple to carry out. For example, if a national household survey has recently been administered, a separate survey of beneficiaries can be fielded using the same questionnaire. The results of this survey can then be combined with those of the national survey to construct treatment and comparison groups through propensity score matching.

3. A good summary of evaluation techniques can also be found in Ravallion (2001).

Still, propensity score matching requires an exhaustive questionnaire to accurately match treated and untreated populations based on their observable characteristics. It demands careful consideration of the extent to which unobserved differences remain between the two samples. Three of the articles in this issue—Chase on Armenia, Newman and others on Bolivia, and Pradhan and Rawlings on Nicaragua—use propensity score matching to estimate impact. In addition, the Armenia evaluation uses a “pipeline” method to minimize possible bias arising from any residual unobserved differences. The logic of this approach is appealing: If there are unmeasured characteristics that make some groups more likely to apply for or receive a social fund project, it should be possible to match outcomes in communities that have already been “treated” with others in the “pipeline”—that is, communities that have self-selected and been preapproved for the next round of interventions.

Instrumental Variables

Two-stage least squares estimation attempts to mimic an experimental design. It relies on a variable, the instrument, which is assumed to be correlated with the probability of treatment but uncorrelated with unobserved determinants of outcomes. Under these circumstances instrumental variables can purge the estimation of treatment effects of such problems as measurement error, reverse causality, or nonrandom program placement. But because instrumental variable estimates are predicated entirely on the validity of the instrument, any correlation between the instrument and unobserved determinants of treatment effects can result in serious biases.

Paxson and Schady, in an evaluation of the Peruvian social fund, use two-stage least squares to estimate the impact of the fund’s investments in school infrastructure on attendance rates. They use the distribution of the progovernment vote as an instrument for social fund spending, building on earlier work showing that changes in the distribution of the vote between 1990 and 1993 affected the distribution of social fund resources in Peru.

V. WHAT ARE THE LESSONS FOR EVALUATION AND DIRECTIONS FOR FUTURE RESEARCH?

The four social fund evaluations in this issue provide plausible estimates of the impact of social fund investments on a number of outcomes, including the availability of infrastructure in communities and the education and health status of beneficiary households. Equally important, they use a range of evaluation techniques to produce those estimates. What lessons do these evaluations of social funds offer for the impact evaluation of social-sector interventions more broadly?

First, when little is known about the likely impact of an intervention and accurately estimating that impact would provide an important public good, randomization is still the most convincing choice of evaluation technique. The random

selection of treatment and control groups provides a solid basis for an impact evaluation even when randomization is on project promotion.

Second, when randomization is not an option, because of resistance to randomization or a desire to reach all eligible beneficiaries, researchers and policymakers should be opportunistic. Credible instrumental variables are hard to come by, but when available they can provide convincing estimates of impact. Propensity score matching is a promising approach to constructing a counterfactual based on similarities between treatment comparison groups. All the evaluations in this issue except the Bolivian case applied propensity score matching to existing data sets (such as household surveys), often supplemented with data collected on social fund beneficiaries—a practical way of reducing the cost of an impact evaluation.

For social funds, two areas of future research seem particularly fertile. An important one is to estimate whether social funds are more (or less) cost-effective than comparable interventions in achieving a particular impact, such as raising enrollment or reducing infant mortality. Such information is indispensable in making the kinds of tradeoffs that policymakers face daily. The evaluations in this issue provide a benchmark against which other interventions could be measured. A second useful exercise would be to estimate the impact of social funds on other outcomes, such as the organizational capacity and social capital of beneficiary communities.

REFERENCES

- Ravallion, Martin. 2001. "The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation." *World Bank Economic Review* 15(1):115–40.
- Rawlings, Laura B., Lynne Sherburne-Benz, and Julie van Domelen. 2002. "Evaluating Social Fund Performance: A Cross-Country Analysis of Community Investments." World Bank, Social Protection Network, Washington, D.C.