



Evaluating Social Fund Impact: A Toolkit for Task Teams and Social Fund Managers

Sarah Adam

October 2006

EVALUATING SOCIAL FUND IMPACT

A Toolkit for Task Teams and Social Fund Managers

Sarah Adam

October 2006

ABSTRACT

In response to the growing demand for evidence of how social fund (and other CDD) projects achieve significant and measurable improvement in peoples' lives, this toolkit aims to increase the number, expand the topical coverage and improve the quality of social fund/CDD impact evaluations. Designed for task teams and managers of social fund and other CDD projects, as distinct from evaluation practitioners, the toolkit first defines impact evaluation and introduces related concepts and methodologies and then guides this audience through the process of planning, designing and implementing an impact evaluation.

ACKNOWLEDGMENTS

This toolkit was prepared under the guidance of Samantha de Silva (HDNSP), David Warren (LACHD), Robert Chase (SDVSC) and Julie van Domelen. Written by Sarah Adam, consultant, HDNSP, it represents the culmination of the efforts of a team of people that includes: Robert Chase, Alexei Monsarret, Maritza Paredes, Laura Rawlings, Lynne Sherburne-Benz, Julie Van Domelen, and David Warren. It draws heavily on manuals for evaluating the impacts of poverty and social programs to introduce impact evaluation methodologies and distills lessons learned from social fund/CDD impact evaluations as cited in the text. Paul Gertler, Ian Walker, Paul Wassenich, Rodrigo Serrano, Mariana Felicio, Anush Bezhanyan, Caroline Mascarell and Hilarian Codippily provided detailed and incisive comments which contributed substantially.

TABLE OF CONTENTS

INTRODUCTION	4
DEFINING IMPACT EVALUATION	6
PLANNING FOR AN IMPACT EVALUATION	15
DESIGNING AN IMPACT EVALUATION	18
Step 1: Decide what to evaluate and delineate indicators	18
Step 2: Explore data availability	21
Step 3: Select methodologies	22
Step 4: Determine data requirements	25
Step 5: Obtain data	28
Step 6: Analyze, report and disseminate data	32
CONTRACTING AN IMPACT EVALUATION	34
CONCLUSION	38
REFERENCES	62

Boxes

1. Monitoring and Impact Evaluation
2. Impact evaluations have improved project performance and influenced policy
3. Areas ripe for further research
4. Measuring impact
5. Impact evaluation resources
6. Recent evaluations study under-researched topics
7. Indicators should be
8. Using a mixed-methods approach
9. Collecting baseline data

Tables

1. Comparison of Quantitative and Qualitative Approaches for Evaluation
2. Linking Data Requirements to Evaluation Methods
3. Indicative Areas of Impact

Annexes

1. Quantitative Impact Evaluation Methodologies
2. Cost-effectiveness Analysis
3. Framing the Objectives
4. Main Data Collection Instruments for Impact Evaluation

INTRODUCTION

As part of an increasing emphasis on community participation in development, the World Bank over the past almost twenty years has financed a rapidly growing number of social funds throughout the world. During that time social fund objectives have expanded from short-term employment and stabilization to longer-term poverty reduction and capacity-building efforts. The institutional designs of social funds and their operating procedures have diversified to achieve these broader objectives. This ability to adapt and innovate is widely recognized as one of the model's greatest strengths.

The scope and scale of the social fund portfolio, heightened interest in community-driven development (CDD) as part of the Bank's core poverty reduction strategy, and the continuing evolution of social fund approaches argue for evaluation of how they (and other CDD projects) achieve significant and measurable improvement in peoples' lives. There is growing demand both from within the Bank and other stakeholders for evidence of actual impact and insights into what works well and what does not in different contexts in order to improve the effectiveness of interventions and allocate resources most effectively.

Recent assessments have called for increasing the quantity and improving the quality of social fund/CDD impact evaluations. They have highlighted the need to focus greater attention on:

(i) extending the focus of impact evaluations to reflect the broader objectives of more recent social funds, including the effects on social capital and community dynamics and on public sector management and institutions; (ii) improving immediately the rigor of evaluation designs; (iii) developing common methodologies and approaches to measuring CDD project performance in order to compare relative performance across programs; and (iv) adding cost analysis to determine whether social funds are more or less cost-effective than comparable interventions in achieving a particular impact.¹

Conducting impact evaluation, however, is challenging. Impact evaluations are technically complicated, often expensive and time-consuming, and constrained by limited availability and quality of data and lack of commitment. Specific characteristics of social funds and other CDD projects, including their objectives, divergent approaches and demand-driven nature, add to the complexity. Yet recent evaluations are employing practical and creative methodologies to accommodate some of these inherent obstacles while maintaining rigorous evaluation standards.

The objective of this toolkit is to assist in the mainstreaming of impact evaluations of social funds and other CDD projects. Recognizing that - like every CDD project - each evaluation is tailored to a unique context, the toolkit seeks to guide task teams and managers of social fund and other CDD projects (as distinct from evaluation practitioners) through the common steps of planning, designing and implementing an

¹ Wassenich and Whiteside (2004), Rawlings, et. al. (2004), Mansuri and Rao (2003), OED (2005, 2002).

impact evaluation. It draws heavily on manuals for evaluating the impacts of poverty and social programs to introduce and apply impact evaluation methodologies, and distills lessons learned from social fund/CDD impact evaluations.

Chapter 1 defines impact evaluation in the broader context of monitoring and evaluation and outlines impact evaluation concepts and techniques. Chapter 2 covers planning for an evaluation. Chapter 3 details steps in designing an impact evaluation. Chapter 4 addresses key issues in contracting. Annexes include technical information.

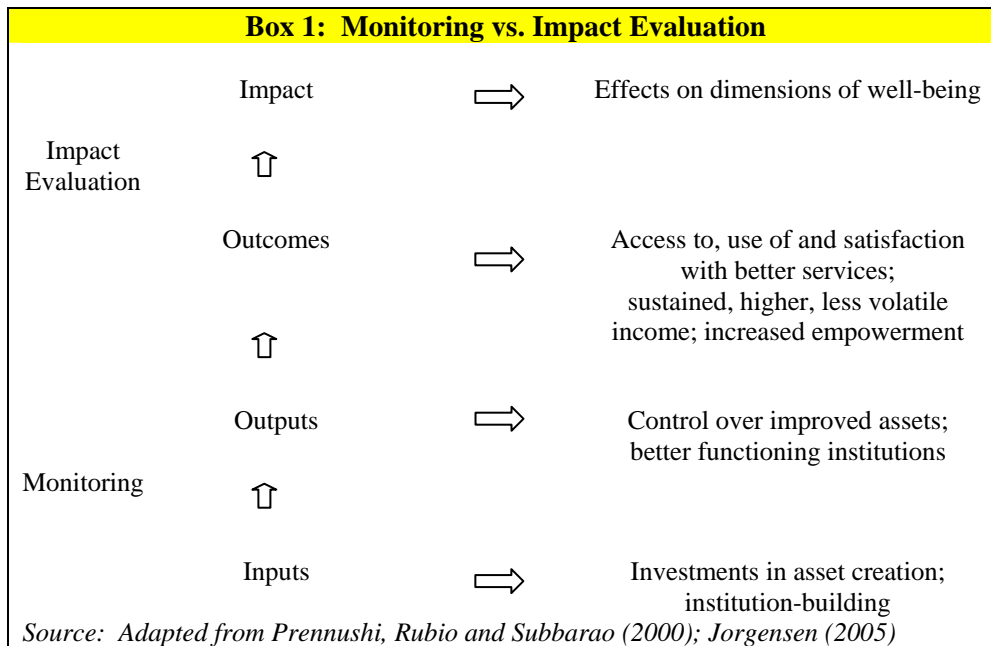
DEFINING IMPACT EVALUATION

Before presenting the steps in planning, designing and implementing an evaluation, this section defines impact evaluation within a comprehensive monitoring and evaluation framework and introduces users to quantitative and qualitative impact evaluation methodologies. Annex 1 describes in more detail the principle quantitative methods and estimation techniques used in impact evaluation.

The goal of monitoring and evaluation is to inform management decisions about whether and how to adjust project design or implementation arrangements to better achieve a project's intended objectives. Monitoring and evaluation results can be used in various ways to improve project performance, including to: (i) clarify the objectives and logic of a project; (ii) reveal impediments to implementation; (iii) prove program integrity and success; (iv) identify lessons for similar projects; (v) inform resource allocation decisions; and (vi) measure project results.

Monitoring and evaluation are different yet complementary components of an integrated system, each of which addresses a distinct aspect of the project. Monitoring involves the continuous and systematic collection and analysis of data on inputs, outputs, processes and assumptions to determine whether a project is being implemented as designed and make appropriate adjustments. Evaluation, broadly defined, is the objective assessment of an ongoing or completed project, including its design, implementation and results, to determine the relevance and achievement of objectives, project efficiency, effectiveness, impact and sustainability. Within evaluation, *process evaluation* examines how the project operates and addresses problems in service delivery and *cost analysis* assesses project costs compared to benefits and to alternative uses of the same resources. *Impact evaluation* seeks to determine whether the project had the desired effects on individuals, households and institutions and whether those effects are attributable to the project.² While monitoring concerns inputs and corresponding outputs, evaluation focuses on impacts and outcomes while considering the causal links to project objectives from outputs and the inputs intended to produce them through project processes and given specific assumptions.

² Kusek and Rist (2004), Baker (2000) and Valadez and Bamberger (1994)



Impact Evaluation

Well-designed and rigorously implemented evaluations are critical for assessing project effectiveness. By addressing such questions as:

- How did the project affect the beneficiaries?
- Were the improvements a direct result of the project (and not of other intervening factors that occurred simultaneously)?
- Did impacts vary across different beneficiary groups, regions or over time?
- Were there any unintended effects of the project?
- How could project design be modified to improve impact?
- How effective is the project compared to alternative interventions?

impact evaluations can enhance project results; inform decisions on whether to expand, modify, replicate or eliminate projects; identify lessons to be adopted in new projects; compare the effectiveness of alternative interventions thereby improving the efficiency of resource allocation; and strengthen the accountability for results.

Box 2: Impact evaluations have improved project performance and influenced policy

In *Honduras*, results of the ex-post evaluation of the Social Investment Fund II (FHIS 2) led to a redesign of the criteria for supporting water systems, including training components; the development of baseline data on incoming projects to facilitate future measurement of impacts; a strengthening of subproject supervision; and efforts to ensure more systematic consultation with beneficiary communities.

In *Nicaragua*, impact evaluation findings of the Emergency Social Investment Fund (FISE) led to the temporary suspension of new sewerage projects while better targeting instruments were designed; the financing of complementary inputs to basic physical works (e.g., access to roads and living quarters for teachers in rural areas); and strengthening of social fund project appraisal and monitoring and evaluation systems.

In *Indonesia*, a study of the impacts of the Kecamatan Development Program (KDP) on household welfare and community organization and perceptions provided useful inputs into future program design, including the redesign of the economic loan portfolio, community outreach materials to improve transparency, and survey methodologies and topical coverage.

In *Tanzania*, based on preliminary results of the impact evaluation of Tanzania Social Action Fund I (TASAF I) the design of the follow-on project was modified to enhance community ownership and reduce the chances of health facilities being constructed where Government lacks sufficient resources to fund recurrent costs. Furthermore, the collection of baseline data strengthens significantly the capacity to evaluate TASAF II.

Source: Rawlings et. al. (2004), Wong (2003), Lenneya (email 5/2006).

Social Funds and Impact Evaluation

Designed as innovative and dynamic delivery systems, social funds as well as other CDD projects require robust monitoring and evaluation systems to ensure effective learning and intended results. Social funds typically have strong monitoring capabilities which use sophisticated computerized management information systems to track whether inputs have resulted in intended outputs. Most funds also routinely carry out studies, including beneficiary assessments which explore beneficiary perceptions of performance at different stages of the subproject cycle, and technical audits which assess the quality of works built. Using these tools, social funds regularly report on indicators such as: the efficiency of subproject processing, sectoral and geographical distribution of subprojects, numbers of beneficiaries trained, technical quality of construction; and levels of beneficiary satisfaction with the process and outputs.

Box 3: Areas ripe for further research

In their study of CDD impact assessments, Wassenich and Whiteside identify knowledge gaps in existing research and recommend attention to:

- gender, conflict and leadership issues;
- project effects on, and interaction with, decentralization processes and public sector management;
- performance of CDD processes compared to more centralized approaches in achieving desired outcomes;
- impact of various components and processes within a CDD project;
- long-term sustainability of welfare impacts and infrastructure outcomes; and
- alignment of projects with *ex-ante* beneficiary priorities.

Source: Wassenich and Whiteside (2004)

The majority of social fund and other CDD projects, however, have not conducted reliable and rigorous impact evaluations, although the number is increasing.³ Overall, the most reliable and extensive group of evidence concerns CDD impacts on primary welfare measures, particularly education and health outcomes.⁴ As a result, much less information is available as to whether through a social fund, the provision of desks increases school attendance and improves test scores; the building of a health center leads to increased use and a decline in child mortality rates; or participation in subproject processing enhances local capacity.

The paucity of impact evaluations is in part due to the fact that conducting impact evaluation is challenging and costly with obstacles complicating most every step of the process. They are technically demanding, sometimes deemed too expensive and time-consuming, and are constrained by limited availability and quality of data and lack of commitment. Furthermore, evaluations are often planned after an intervention has already begun which limits the methodologies available.

Specific characteristics of social funds and other CDD projects add to the complexity. The fact that communities elect to participate complicates evaluation design: it makes establishing random program placement or fielding a baseline survey in a sufficient number of future participant and comparison areas difficult; and introduces bias based on unobserved differences between similar communities that do and do not apply that is technically challenging to correct. Critical objectives such as building social capital and empowering communities are hard to measure. In addition, they depend heavily on context and thus extracting lessons or measurement tools from one place to another requires great care.⁵

³ Wassenich and Whiteside (2004)

⁴ See Rawlings, et. al. (2004)

⁵ Rawlings et. al. (2004); Wassenich and Whiteside (2004)

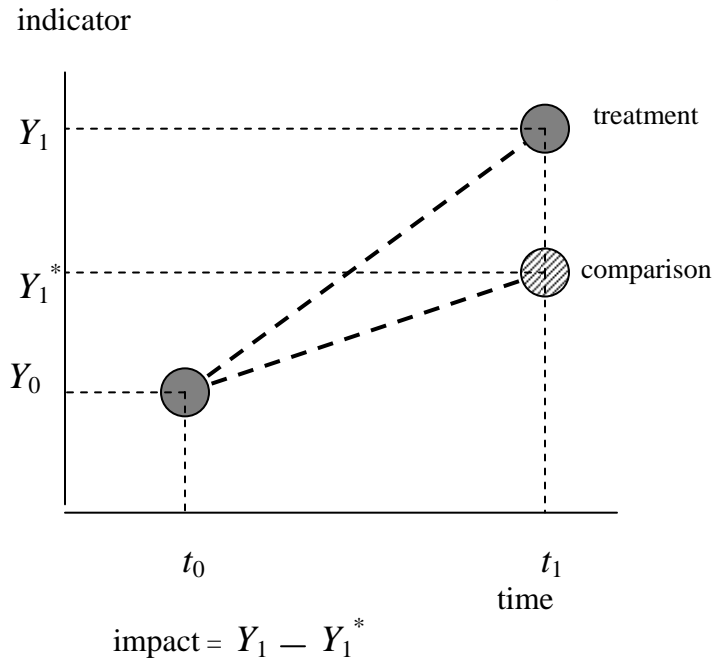
Despite these challenges, evaluations are being designed which accommodate some of the obstacles inherent to demand-driven projects while maintaining rigorous evaluation standards, and promising approaches for evaluating the impact of social fund and other CDD projects are emerging.⁶

Impact Evaluation Methodologies

To estimate the magnitude of effects and convincingly establish a causal link to a project, impact evaluation relies on the construction of a *counterfactual* – an attempt to estimate what a given outcome would have been for project beneficiaries if the project had not been implemented. The counterfactual outcome is never actually observed as people cannot simultaneously participate and not participate in a project.

In order to model the counterfactual, impact evaluation establishes a control or comparison group (those who do not participate or receive benefits) and compares it with the treatment group (those who do receive the intervention). Quantitative methodologies for evaluating impact fall into two broad categories: (i) *experimental* or randomized designs; and (ii) *quasi-experimental* or nonrandomized designs. In both cases, the comparison group consists of individuals, households, communities or other units that possess similar characteristics to those in the treatment group. By comparing the average of the changes in the treatment group to that of the comparison group, the evaluation can determine how great a change occurred and whether the change was due to the intervention or to some other factor (e.g. an improvement in the overall economy or another government program). Without a counterfactual analysis there is no way to know if the observed outcomes are in fact attributable to the project.

⁶ Chase presentation (2005)

Box 4: Measuring impact

Source: Ravillion presentation (2005)

Constructing an appropriate comparison group is technically complex. A main challenge is addressing bias that stems from differences between participants and non-participants. Two types of bias can distort the results of an evaluation: (i) observable and (ii) unobservable. *Observable* factors can be measured using data collection instruments and then accounted for when measuring impacts. Examples include the criteria used to target beneficiaries (e.g., geographic location, level of poverty, social infrastructure conditions) and education levels. *Unobservable* characteristics are either not known by the researcher or are not easily measured (e.g., communities' motivation, social capital, organizational capacity). With no way to directly observe such traits, it is difficult to construct an accurate comparison group. Bias can result in either project impacts being overestimated (upward bias) or underestimated (downward bias).

Establishing a control or comparison group allows for a “with or without comparison”. Beneficiary groups can also be compared before and after the intervention if baseline and follow-up data is available. Since a “before and after” method uses information from the same population, beneficiaries' predispositions and innate abilities are constant. Such a comparison does not, however, avoid the influences of exogenous factors that affect the beneficiaries (e.g., geographical catastrophes, macroeconomic shocks and national development programs) that distort the measurement of social fund outcomes. Baseline data, however, may be crucial to reconstructing why certain events took place and controlling for these events. Combining baseline data with comparison group-based methods allows evaluators to verify the integrity of treatment and comparison groups, and thus greatly increases the credibility of the evaluation.

Quantitative research is empirically rigorous, impartial and objective. Conducted appropriately, quantitative approaches to evaluation can: (i) permit generalizations to be made about larger populations on the basis of much smaller, representative samples; (ii) assess causality of the impacts of given variables on project outcomes; and (iii) allow other researchers to validate the original findings by independently replicating the analysis. Quantitative techniques, however, have limitations. Many important issues, such as those related to process, cannot be meaningfully quantified or adequately understood without reference to context. Quantitative surveys tend to reflect the preconceptions and biases of their authors and limit opportunities for new or unexpected findings.⁷

Annex 1 describes the main quantitative methodologies and estimation techniques used in impact evaluation and their advantages and constraints.

Qualitative techniques

Qualitative methods focus on understanding processes, behaviors and conditions as perceived by the individuals or groups studied and can also be used to assess impact. They often yield critical insights into beneficiaries' perspectives, the processes and context that may have affected outcomes and a more nuanced interpretation of results than is observed in quantitative analysis.

Qualitative techniques, which use relatively open-ended methods to collect data, include:

- *Conversational interviews* with key stakeholders;
- *Focus group* discussions with intended beneficiaries;
- *Group meetings* open to all community members.
- *Direct observation* in targeted communities; and
- *Case studies* which collect descriptive or explanatory information.

These research tools are often complemented by other survey techniques ranging from traditional questionnaires to those used in *participatory rural appraisal* (e.g., mapping, ranking and trend analysis) which rely on participants' knowledge of the conditions surrounding the project being evaluated.

Rather than using statistical analysis and estimation methods, qualitative methodologies often verify the validity and reliability of data through triangulation. Triangulation involves the comparison of data sources, collection methods, investigators and theory to determine the degree of likely biases and the reliability of the information gathered.

The advantages of qualitative methodologies are that they: (i) can be tailored to the needs of the evaluation; (ii) allow in-depth study of selected issues, cases or events; (iii) can be carried out quickly by using rapid techniques; and (iv) provide critical insights into beneficiary perspectives, the dynamics of a particular reform and the conditions and processes that may have affected project impact.

⁷ Rao and Woolcock (2003)

There are, however, significant disadvantages to relying on qualitative techniques alone. These limitations include: (i) subjectivity in data collection as researchers make interpretive judgments; (ii) dependence of the validity and reliability of data on the skill, sensitivity and training of the researcher; (iii) lack of statistical robustness given typically small sizes which makes it difficult to generalize findings to a larger population; (iv) difficulty in determining the causality of project impact; and (v) difficulties in replicating and thus independently verifying results.

Integrating Quantitative and Qualitative Techniques⁸

Because measuring the counterfactual is at the core of impact evaluation, qualitative techniques are commonly used with quantitative methodologies. While they differ in approach (see Table 1), the strengths and weaknesses of qualitative and quantitative approaches potentially complement each other. Combining both methods can provide a more comprehensive and insightful impact evaluation than either approach could produce on its own.

Table 1: Comparison of Quantitative and Qualitative Approaches for Evaluation

<i>Aspect</i>	<i>Quantitative approach</i>	<i>Qualitative approach</i>
Objectives	<ul style="list-style-type: none"> To assess causality and reach conclusions that can be generalized 	<ul style="list-style-type: none"> To understand processes, behaviors, and conditions as perceived by the groups or individuals being studied
Data collection instrument	<ul style="list-style-type: none"> Structured, formal, pre-designed questionnaires 	<ul style="list-style-type: none"> In-depth, open-ended interviews Direct observation Written documents (e.g., open-ended items on questionnaires, personal diaries, records)
Sampling Methodology for analysis	<ul style="list-style-type: none"> Probability sampling Predominantly statistical analysis 	<ul style="list-style-type: none"> Purposive sampling Triangulation Systematic content analysis Gradual aggregation data based on selected themes

Source: Prennushi, Rubio and Subbarao (2000)

Skillfully integrating quantitative and qualitative techniques can greatly increase the richness of an impact evaluation by providing credible estimates of impact as well as an explanation of the processes and interventions that yielded project outcomes. Although quantitative and qualitative evaluations can proceed in parallel with findings compared and combined during analysis, the greatest value-added from using the two methods is realized through an iterative approach where preliminary results from one type of data collection are explicitly intended to inform the design of a future round of data collection. For example, early qualitative work, with the help of quantitative data collection techniques, can be used to analyze and understand the social, economic and political context in which the project takes place; assess project processes; and explore research topics. Hypotheses derived from these field findings can be used to design the survey

⁸ Rao and Woolcock (2003)

and sample to inform the extent to which the results can be generalized to a larger population.

Combining methodologies, however, does present difficulties:

- Integrating different perspectives requires individuals with different skill sets, which increases the costs of an evaluation in terms of time, talent and resources;
- Coordinating large teams of people with diverse backgrounds generates additional coordination challenges; and
- There is little (though evolving) evidence regarding how to best combine different methods under which circumstances.

Despite these challenges, social fund and other CDD projects make ideal candidates for a mixed-methods approach given the critical importance of process and context and other less tangible and easily quantifiable factors in assessing and understanding outcomes.

Cost-effectiveness Analysis

Determining whether a project is more (or less) cost-effective than comparable interventions in delivering services and/or achieving a specific impact is extremely important, particularly when resources are limited. Although not strictly concerned with assessing impact, cost-effectiveness analysis greatly enhances the policy implications of impact evaluation and therefore its inclusion is recommended.⁹

Cost-effectiveness analysis involves calculating all project costs (in monetary terms) and comparing them to benefits (non-monetary outputs, e.g., number of units delivered) to compute a cost-effectiveness ratio – the value of cost per unit of benefit. This ratio is then compared across interventions to determine relative efficiency. Efficiency can be assessed further by comparing overhead costs, speed of execution and the amount of community resources leveraged across different delivery mechanisms.

As with impact evaluation techniques, actual application of cost-effectiveness analysis is complicated. It is difficult to find comparators, quantify costs and benefits, and ensure that ratios are comparable across institutions. Annex 2 provides more detail on implementing cost-effectiveness analysis.

⁹ Baker (2000)

PLANNING FOR AN IMPACT EVALUATION

Before an evaluation is designed, it is first necessary to lay the groundwork. This chapter details ingredients important to the success of an impact evaluation as well as the parameters that frame an evaluation. Preparing for the evaluation is primarily the responsibility of social fund/CDD project management.

Impact evaluations demand substantial time, resources, and analytical capacity. In considering whether to conduct an impact evaluation, such factors should be weighed against the value of the information the evaluation is expected to yield. In general, the costs of an impact evaluation are likely justified if the intervention: is innovative and new insights can be generated, is replicable, is of strategic relevance for reducing poverty, helps fill knowledge gaps about cost-effective poverty reduction alternatives, and/or the policy impact of the results is likely to be substantial. Most social fund and other CDD projects - or aspects of their operations - meet these conditions.¹⁰

Ingredients for success

Impact evaluations are sometimes viewed by governments as taxes imposed by the lending institution, to be used for academic purposes with limited operational impact. To help ensure an evaluation's relevance and usefulness, managers should:

Garner political and financial support. Evaluations need to be carried out with strong links to decision-makers and the decision-making process if they are to have an impact on policy. Program administrators, government policymakers and financiers must be convinced that the questions to be addressed are relevant to decisions about whether to expand, modify or eliminate the project, and that the evaluation design is rigorous and credible. In order for evaluations to be worth foregoing investments in communities, studies need to focus on issues these officials consider priorities.

Plan early. Impact evaluation gets easier, less costly and more credible the earlier in the project cycle that it starts. Ideally, planning and designing an impact evaluation should start during project preparation. Early and careful planning, as detailed in the following chapter, helps to focus project goals, allows the collection of baseline data including the testing and refinement of survey instruments, and expands the range of methodological options open to evaluators thereby enhancing the quality and relevance of the results.¹¹ Planning also involves including a realistic monitoring and evaluation budget from the start of the project. (See the following section for a discussion of financing sources).

Understand the project and its context. The evaluation team must gain a thorough understanding of the intervention and the context in which it takes place, including how the project's inputs and processes given relevant assumptions are supposed to achieve specific outputs and outcomes, in order to accurately tailor the evaluation's design. Such knowledge is gained through careful reading of relevant project documents, but more

¹⁰ As the knowledge base of social fund/CDD impacts grows, a strong case may be made for limiting and focusing future evaluations.

¹¹ Goldstein presentation (2005)

importantly, through discussions with project staff, beneficiaries and other stakeholders to understand how a project actually operates. Evaluators should also be well-informed about what information is routinely collected during each stage of the subproject cycle as well as available sources of data outside of the project (see Chapter 3 Step 2).

Involve stakeholders. Involving all stakeholders in the design and implementation of impact evaluations is necessary to ensure that the process is transparent and credible and that the results will provide information relevant to their needs. Stakeholders include project beneficiaries, social fund managers and staff, government policy makers and financiers, public agencies, non-governmental organizations, bilateral donors and international development institutions.

Coordination among donors can facilitate the design and financing of the evaluation as well as the use of the results. Coordination is also important to avoid duplication of evaluation efforts, to rationalize the financing of evaluations, and to foster collaboration in supporting the implementation of policy changes suggested by evaluation results.

Dimensioning the evaluation

The planning process also involves determining the scale and complexity of the evaluation design. As early as possible in the planning process, project teams need to address the following questions: When is the data needed? How much will an evaluation cost? How will the evaluation be financed? What is the in-country capacity to conduct impact evaluation?

Timeframe and Cost. The timeframe and cost of an impact evaluation depend largely on: (i) the depth and breadth of the evaluation; (ii) its design; and (iii) the availability of existing data. Impact evaluations can take months or years to conduct. Random assignments and evaluations using baseline and follow-up data take longer than ex-post comparisons (but enhance quality). The existence of relevant data and/or routine data collection can significantly reduce the time (and cost) of an evaluation. Regardless of the evaluation methodology, however, sufficient time has to have elapsed for the intervention to produce measurable results. This time lag presents significant logistical and financing complications as impact evaluations often need to be conducted after a project has been completed. By that time, support for an evaluation may have waned and project financing may not be available.

The costs of evaluations vary widely. Expenditures on social fund impact evaluations have ranged from approximately 0.25 to 2.0 percent of total project cost. Data collection (including instrument design, sampling, fieldwork, and data management) typically accounts for the largest share of expenditures (approximately one-third) followed by evaluation design and analysis activities carried out by local and international consultants. Data collection costs are higher when new survey instruments must be designed and where data is collected as part of the evaluation effort rather than building on existing household surveys.

Financing. Evaluations can be financed by a project, other government resources, grants, outside donors or some combination of these sources. In the case of social funds, countries have absorbed the majority of the cost, but with substantial resources coming from World Bank research and bilateral donor grants and the World Bank's overhead budget. Funding through project loans gives governments control over the evaluation contracting process which may increase the likelihood that the results will inform policy yet may also compromise data quality. Evaluations which are entirely externally funded with Bank staff in charge may be more rigorous, but risk limiting the relevance and impact of the results. Given that the lessons generated by impact evaluations are public goods valued by stakeholders beyond those working on the project being evaluated, a strong case can be made for countries to try to secure technical and financial resources from development agencies.¹²

Capacity. Coordinating and executing an impact evaluation is technically challenging, requiring skilled staff throughout the process. Capacity varies greatly among countries and there are trade-offs between contracting international technical assistance and using local capacity. While the former may result in a more rapid and possibly less costly assessment, the latter will help build capacity in the country – often an important evaluation goal.

Given heightened interest in impact evaluation within the World Bank and by the broader donor/lender community, many resources exist to help project teams in preparing for, designing, implementing and financing impact evaluations. Box 5 details some of these resources available.

Box 5: Impact evaluation resources

Evaluation specialists: World Bank resource persons to advise on the design of and peer review impact evaluations can be found in most regions, sectors, social fund and CDD anchors, and DEC. The internal PovertyNet website provides access to a roster of “external” impact evaluation experts.

Impact evaluation clinics: the office of the Chief Economist of the HD Network and the Thematic Group on Poverty Analysis, Monitoring and Evaluation offer on-demand impact evaluation training clinics and field-based evaluation training to support task teams in the design of impact evaluations. The Africa region conducts impact evaluation workshops bringing together task teams, government counterparts and local and international researchers to learn about impact evaluation and apply it to the design of impact evaluation for their specific projects.

Financing – sources of financing include project loans, bilateral donors, trust funds, Bank budget for Bank staff, and meta-evaluations on strategic themes are being conducted under the Development Impact Evaluation (DiME) initiative. There is a new AAA product line for impact evaluations (must be funded by Bank resources or trust funds).

Websites – The Thematic Group on Poverty Analysis, Monitoring and Evaluation, managed by PREM, hosts the PovertyNet website (www.worldbank.org/poverty click on impact evaluation) which details impact evaluation methods and techniques, data and tools, key readings and links to non-Bank impact evaluation resources. The CDD anchor's website (www.worldbank.org/cdd click on design and implementation) explains the status of impact evaluations of CDD projects and contains links to key resources. The Social Funds website includes social fund impact evaluations and beneficiary assessments.

DESIGNING AN IMPACT EVALUATION

Like every social fund/CDD project, each evaluation must be tailored to its unique context. This section describes the steps common to conducting all impact evaluations. While project teams (Bank task teams and their social fund/CDD project counterparts) determine evaluation objectives, the steps that follow are the responsibility of those contracted to conduct the evaluation. Given the toolkit audience, the level of detail provided is designed to give project team's sufficient knowledge to manage the impact evaluation process.

The goal of project teams, with specialists contracted to conduct the evaluation, should be to design the most robust impact assessment possible given the particular project theory, context and constraints faced. Impact evaluation designs vary in focus, time, cost, feasibility, and the degree of clarity and validity of results. Regardless of the specific design, the following key steps are common to conducting all impact evaluations.¹³

- i. decide what to evaluate and delineate indicators;
- ii. explore data availability;
- iii. select a methodology;
- iv. determine data requirements;
- v. obtain data; and
- vi. analyze, report and disseminate data.

Step 1: Decide what to evaluate and delineate indicators

The first and one of the most important and difficult steps in designing an impact evaluation is deciding what to evaluate. Clear evaluation objectives are essential for setting measurable indicators, identifying data requirements and formulating a coherent evaluation strategy, and thus contribute significantly to an evaluation's success.¹⁴

The logical framework (or logframe) is one analytical tool for identifying and ordering project objectives, key indicators and the information needs around which an evaluation should be constructed. Based on a simple matrix, the logframe lays out the project's results chain by linking information on project objectives with how performance will be tracked, the impacts and outcomes of outputs and how they will be measured, and how inputs are used to deliver outputs.

Evaluation design should start first by understanding the logic of the project (as it is actually being implemented not as originally designed), including the assumptions upon which the project is based. This project theory provides the framework for establishing evaluation goals, objectives and indicators. Given that for many social fund/CDD operations the project processes are critical to the internal project logic, the evaluation should measure outcomes, outputs, inputs and assumptions as well as impacts.¹⁵

¹³ Adapted from Baker (2000)

¹⁴ Baker (2000)

¹⁵ Jorgensen (2005) and Woolcock et. al. presentation (2004)

Study questions must also identify the comparison group that best represents the counterfactual. Communities benefiting from a social fund or other CDD project can be compared to those who have not (or have not yet received) social fund financing, and/or those who have received similar investments through different service delivery systems such as central government line ministries, local governments, or non-governmental organizations. Both comparators are valid, and require different methodological approaches.

In response to budget, time or other constraints, impact evaluations often focus on a few key objectives. Several factors should be considered and weighed in narrowing the range of topics studied, including: the research priorities of government and other stakeholders, the concentration of project resources; the availability and quality of existing information on the impacts of interventions; and gaps in global knowledge about the effectiveness of particular interventions in specific contexts.

The most widely disseminated social fund impact evaluations cover those whose main objective was infrastructure development, and thus these focused primarily on the socio-economic impacts of physical investments.¹⁶ More recent and ongoing evaluations are expanding their focus to capture both the innovations in social fund design and implementation as well as the impacts of a more relevant set of objectives.

¹⁶ see Rawlings et. al. (2004)

Box 6: Recent evaluations study under-researched topics

Corruption in Indonesia: A study of village road projects in Indonesia financed under the Kecamatan Development Project (KDP) examined alternative approaches to fighting corruption. Olken first obtained a direct measure of corruption by comparing independent estimates of what each road actually cost to build with what a village reported it spent. Then using a randomized, controlled field experiment in over 600 villages, he compared the effects of two strategies: top-down monitoring by government auditors and bottom-up participation in the village monitoring process. The study found considerable evidence that increasing the probability of external audits substantially reduced the amount of theft in the project (by about 8 percent). Furthermore, cost-benefit analysis revealed that the benefits from such audits exceed their costs. In contrast, increasing community participation in monitoring shifted the method of corruption from overstated wage payments to overstated expenditures on materials, but had minimal effect on the overall level of corruption.

Comparative effectiveness in Pakistan: The evaluation of the Pakistan National Rural Support Project (NRSP) conducted by the Bank's Development Economics group (DEC) will explore both the effectiveness of the community driven-development approach employed by NRSP compared to alternative interventions, and the efficacy of different delivery mechanisms within the project. Based on a sample of 140 villages with about 450 comparable government projects, the study will assess whether community infrastructure created under the project is of better quality, more cost-effective and more sustainable (based on design, construction, cost and maintenance) than similar infrastructure (in size and type) created under the government's more traditional "top-down" approach. The evaluation also seeks to examine whether an increase in social mobilization efforts enhances the capacity for collective action in target communities. From a sample of 75 randomly selected villages, one-third will serve as the control group and will receive no intervention, in another one-third of the villages NRSP will deliver the program at its current average level of coverage, and in the remaining villages the project will intensify its social mobilization efforts. By varying the intensity of the coverage, the evaluation will examine whether NRSP's objectives are more likely to be achieved when scaling up the project by increasing the depth of coverage within villages compared to expanding coverage to more villages. The DEC study further seeks to assess the impact of training delivered as part of its micro-credit program by randomly selecting 80 NRSP villages and randomly assigning them to receive training or no training. The results will help answer whether lack of information (on profitable enterprises, marketing opportunities or business management) constrains the uptake or effective use of credit.

Sources: Olken (2005); Wassenich and Whiteside (2004); Concept Note for AAA Proposal for the Evaluation of NRSP (2006)

Once the critical research areas are clearly defined, project teams in consultation with key stakeholders need to develop indicators. Box 7 lists common features of good indicators.

Box 7: Indicators should be:

Relevant – measuring factors that affect the projects;

Selective – few and meaningful;

Clear – precisely defined so that their measurement and interpretation is unambiguous (more – or less – is unmistakably better);

Specific – reflecting aspects the project intends to change and avoiding measures largely subject to external influences, and not easily able to be manipulated to show achievement where none exists;

Collectable – feasible and (if not already available) relatively inexpensive to collect within a reasonable timeframe;

Variable – across areas, groups and over time and *sensitive* changes in policies, programs institutions.

Source: Adapted from Prennushi, Rubio and Subbarao (2000)

In selecting indicators, it is important to understand the timeframe within which changes can be expected. In cases where the full effects of an intervention will only be realized in the longer term, indicators of intermediate outcomes can judge the direction and speed of realizing the project's objectives (and should reveal opportunities for improvements in project design).

All stakeholders – beneficiaries, social fund field, regional office and technical department staff and managers, governments, non-government organizations, and donors – share an interest in the evaluation's objectives and indicators, and thus should participate in their selection.

In order to help project teams to clarify evaluation objectives and define indicators, Annex 2 provides a framework for categorizing broad areas of inquiry relevant to most social funds and many other CDD projects.

Many social fund impacts are directly in line with the Millennium Development Goals (MDGs), including income gains, increases in primary school enrollment and school achievement, reductions in infant and child mortality, and expanded access to water and sanitation. Impact indicators and data for each evaluation should track those MDGs relevant to each particular subproject.

Given time, cost and capacity bounds, the process of selecting objectives and indicators should include an assessment of what data is available, what data is routinely collected, and what new information can be realistically collected within these constraints.

Step 2: Explore data availability

Impact evaluations use many types of data, ranging from quantitative panel data to qualitative information gleaned from informal, unstructured interviews. As data collection can be both time-consuming and expensive, evaluators should first take stock of existing data (coverage and quality) and data sources and plan additional data collection to maximize its usefulness. Existing data on project impacts can come from the project itself and from external sources.

The social fund may have:

- information on beneficiaries (and non-beneficiaries, if available) which will help determine sample size, construct the sampling frame, and select the sample;
- administrative data on geographic distribution, expenditures and composition of subprojects;
- poverty map data;
- monitoring and output data from the management information system;¹⁷

¹⁷ An important lesson from impact evaluations to date for project design is that social funds should routinely collect baseline data from subprojects as part of the appraisal and follow up data from representative samples of subprojects to monitor changes that occur once the investment is implemented. These efforts would provide much of the basic data needed to routinely evaluate social funds' progress. (Rawlings et. al., 2004)

- results of past evaluations, surveys, audits and beneficiary assessments; and
- administrative cost data needed for cost-effectiveness analysis.

External data sources to be examined include:

- household income or expenditure surveys;
- Living Standards Measurement Studies;
- censuses;
- school records (*e.g.*, attendance, repetition, examination performance);
- public health records (*e.g.*, infant mortality, infectious disease incidence);
- specialized surveys conducted by universities, NGOs or consulting firms; and
- monitoring data from government-administered programs.

Taking advantage of existing data or piggybacking on ongoing or planned household surveys is often more operationally feasible and cost-effective than carrying out a separate survey. Identifying treatment villages in a general household survey that measures required indicators can be used in techniques such as propensity score matching (provided the survey sample and treatment village coverage are large). If a survey is planned that will cover the population of interest, the evaluation may be able to introduce a series of questions or add a qualitative survey to supplement the quantitative data. Yet the savings gained in piggybacking on existing surveys should be weighed against potential problems arising from the timing of the data collection effort, the lack of flexibility of the questionnaire design, and the quality of fieldwork and data processing.

Step 3: Select methodologies

“The art of good evaluation is to draw eclectically from the whole menu of methods to find the most cost-effective combination appropriate to each setting.”¹⁸

Once evaluation objectives have been identified and existing data sources explored, choices about methodology can be made. Each impact evaluation will have unique characteristics requiring different methodological approaches. As described in Chapter 1 and Annex 1, there are a variety of techniques for evaluating project impact – none is ideal and each has advantages and disadvantages. In all cases, using multiple quantitative methods combined with qualitative techniques enables triangulation and significantly strengthens evaluation results.

The potential variety of methodologies open to evaluators depends on the evaluation questions, time and resource constraints, data availability and the desired level of results. Evaluation questions should be carefully matched with appropriate and feasible quantitative and qualitative methodologies.¹⁹

¹⁸ Ravallion as quoted in Wassenich and Whiteside (2004)

¹⁹ Standard methodologies for evaluating interventions in specific sectors can usually be adapted to fit certain evaluation questions.

The choice of methodologies also depends on the timing of the evaluation design relative to project investment decisions. The earlier an evaluation is planned, the more methodological options – particularly among quantitative techniques – are available. If the evaluation is designed during preparation of the project (as advised) and before subprojects have been selected, then both experimental and quasi-experimental techniques to model the counterfactual are open to evaluators. If social fund/CDD subprojects have not yet begun implementation or implementation is phased, collecting baseline data is strongly recommended. (See Step 4 for further discussion of the value of baseline data).

If the evaluation is designed during subproject implementation or after subproject completion and only post-intervention data is available, the choice of quantitative evaluation techniques is limited. Quasi-experimental designs, however, can provide a sufficiently rigorous evaluation while remaining flexible in their requirements and applicability. The data requirements for these methods are more varied, which can (although do not necessarily) lower data collection and analysis costs.

As noted above, the most rigorous impact evaluations use more than one quantitative technique – both combined and separately – but compared – to validate the robustness of the assumptions underlying different estimates and thus the findings. For example, baseline and follow-up data can improve the reliability of matching estimates by further controlling for time-invariant characteristics. Field visits can verify the similarity of matched pairs.²⁰

Qualitative techniques

As with quantitative methods, qualitative techniques are best combined to obtain the fullest understanding of topics probed. Semi-structured conversational interviews are often complemented by focus group discussions, participant observation and structured questionnaires. Opinions should be solicited from multiple perspectives – by selecting participants in communities from a mix of occupation, gender, age and wealth and status groups and from different institutions (e.g., beneficiaries, intermediary implementing organizations, local governments). As discussed in Chapter 1 and highlighted in Box 8, the better the integration of qualitative and quantitative techniques, the more complete and convincing the evaluation results.

²⁰ Wassenich and Whiteside (2004)

Box 8: Using a mixed-methods approach

Designed and conducted by a team of Bank specialists and Thai researchers, the *Thailand Social Capital Evaluation* combined quantitative and qualitative evaluation techniques, iteratively, to assess the impact of the Thailand Social Investment Fund (SIF) on village social capital. Lacking baseline data, the team adopted a creative and pragmatic approach to address this common methodological challenge.

Thai academics and practitioners developed a conceptual framework and indicators for measuring social capital. Using existing household survey data, 72 sample villages that participated in the SIF were each matched with six potential comparison villages in the same province using propensity scores. Qualitative researchers, recognized as leading investigators of village social characteristics, visited these potential comparison villages to pair each treatment village with its most similar match based on indicators such as distance to nearest town, access to infrastructure and out-migration levels using data from various government sources combined with interviews with local officials and community organizations. After three days in each of the 144 matched villages, researchers scored each social capital variable on a scale from one to five (enabling quantification of qualitative observations). Analysts then used differences in means and regressions on each indicator to establish significant differences between those villages that participated in the SIF and those that did not. This analysis revealed several significant differences, and field researchers were asked to separate whether observed differences were due to SIF processes selecting villages already well-endowed with certain social capital characteristics (selection effects) or to social fund activities having an impact on those characteristics (impact effects). The mixed quantitative and qualitative methodology produced nuanced conclusions about how SIF both selected villages better endowed with cooperative norms and had an impact on social capital indicators, including enhancing leadership, strengthening networks and increasing villagers' ability to exercise voice to formal authorities.

Source: Chase (2006)

Adapting to constraints

The most rigorous and credible impact evaluations include the following components:²¹

1. comparison or control groups;
2. collection of baseline data; and
3. mixed quantitative and qualitative techniques.

Yet evaluations are often conducted in resource-constrained environments which require compromises. Evaluators should aim to design and carry out the most robust and useful evaluation possible given the particular context and set of constraints.

There are several ways to adjust to cost, time, and capacity limits without fatally compromising the rigor of the methodology and thus the validity of the results. The most effective approach is to scale down a robust evaluation design either by narrowing the ranges of topics covered or reducing the scope of fieldwork while maintaining a sound sampling framework and evaluation methodology.²²

²¹ Wassenich and Whiteside (2004)

²² Any narrowing of evaluation objectives must ensure that individual project needs and the expectations of project administrators and policy makers are met. See Wassenich and Whiteside (2004). For more detailed information on addressing constraints see Bamberger presentation (July, 2005).

Evaluation design can, and often does, evolve between baseline and follow-up stages as projects are often not carried out as planned. Even after the evaluation design has been determined and built into the project, evaluators should be prepared to be flexible and make modifications to the design as the project is implemented.

Step 4: Determine data requirements

Data quality is an important factor affecting the validity of evaluation results. Three main steps assist in determining and ordering data requirements: (i) establishing the unit of analysis and its boundaries; (ii) linking data requirements to evaluation methods; and (iii) preparing a matrix of analysis.

Establish the unit of analysis and its boundaries

The unit of analysis is the entity (e.g., individuals, groups, households, geographical units, social interactions and institutions) that will be analyzed in order to observe the established indicators. The unit of analysis has enormous implications for the collection of data and for the design of data collection instruments. Evaluations with many units of analysis tend to be more complex and expensive. If the data is not already available, researchers must design instruments for each unit of analysis.

The unit of analysis should be delimited as much as possible. For example, if households are selected as a unit of analysis, the evaluation design should establish whether it will cover all beneficiaries' households, or only some of them (e.g., those that received a certain type of intervention, received it for a certain period, belonged to some geographic area such as urban or rural). Boundaries of units of analysis are commonly delimited using subproject type, evaluation period, and/or geographic area.

Link data requirements to evaluation methods

Since the validity of evaluation results depends largely on the adequacy and reliability of data, it is important to use different sources of data collected through quantitative and qualitative methods. While quantitative methods are often employed to collect quantitative data and qualitative methods to obtain qualitative data, quantitative methods can also be used to collect qualitative data (e.g., open-ended response questions included in a large survey) and quantitative data can be derived from a large number of qualitative responses.²³ As with evaluation methodologies, integrating such data and methods (and analysis) results in more convincing recommendations.

Data needs depend on the outcomes to be measured and the evaluation design. Since projects evaluated will cover a range of indicators and require different evaluation designs, data requirements will also differ. Table 2 links the basic evaluation methodologies with data requirements. With most of these methodologies, qualitative techniques can - and should - be used to help delineate indicators, design the survey

²³ Rao and Woolcock (2003)

instrument, and identify controls and variables used for matching, or instrumental variables.

Table 2: Linking Data Requirements to Evaluation Methods

<i>Method</i>	<i>Data requirement</i>		<i>Use of qualitative approach</i>
	<i>Minimal</i>	<i>Ideal</i>	
Experimental	Single project cross-section of treatment and comparison group	Panel data on both treatment and control group. (This allows for difference-in-difference estimation).	<ul style="list-style-type: none"> • Inform design of survey instrument, sampling • Identify indicators
Quasi-experimental			<ul style="list-style-type: none"> • Data collection and recording using: textual data; informal or semi-structured interviews; focus groups or community meetings; direct observation; participatory methods; photographs
Matching comparison	National survey (census, budget or LSMS-type) and over-sampling of project participants	National survey and smaller project-based household survey, both with two points in time to control for contemporaneous events	
Reflexive comparison and double difference	Baseline and follow-up on project participants	Time series or panel on beneficiaries and comparable non-beneficiaries	
Statistical control or instrumental variable	Cross-section data representative of beneficiary population with corresponding instrumental variables	Cross-section and time series representative of both the beneficiary and non-beneficiary population with corresponding instrumental variables	<ul style="list-style-type: none"> • Triangulation • Data analysis

Source: Adapted from Prennushi, Rubio and Subbarao (2000) and Baker (2000)

Baseline data

As noted in Step 3, baseline data on the pre-intervention state of beneficiaries and comparators should be collected prior to project implementation (or a phase of implementation) whenever possible as it can significantly strengthen evaluation results. Combined with subsequent follow-up data, it enables difference in difference estimates of impact (and, at a minimum, reflexive comparisons). Baseline data allows evaluator to verify the integrity of treatment and comparison groups and reconstruct why certain events took place and control for them. Baseline data can also improve project design. Quantitative and qualitative baseline surveys were conducted as part of the KALAHI-CIDSS project in the Philippines as described in Box 9.

Gathering baseline data presents challenges, including gaining clear agreement with the implementing agency on targeting procedures, indicators and sampling; timing the survey if it will delay project launch; financing the survey; and data and techniques available to match treatment and comparison areas.²⁴

Box 9: Collecting baseline data

The KALAHI-CIDSS project in the Philippines aims to empower communities through their enhanced participation in governance and project implementation. In order to help determine project impact through rigorous evaluation and to improve project performance, evaluators conducted quantitative and qualitative baseline surveys in a representative sample of treatment and matched comparison groups. Although the project was already underway in some municipalities, evaluators capitalized on the project's phased approach by implementing the surveys in areas selected for future treatment.

In the fall of 2003, researchers conducted the first round of the quantitative survey that will track 2,400 households and 132 villages before, during and after project implementation. In line with project objectives, the survey collected data on multiple dimensions of poverty, and social capital, empowerment and governance indicators in treatment and comparison communities. To gain a deeper understanding of the findings of the quantitative survey, examine in more detail local-level dynamics and context, and capture the nuances of the actual situation of communities, researchers conducted a qualitative baseline survey in 2005 in selected municipalities participating in the quantitative survey. Researchers used focus group discussions and key informant interviews to gather information from women, men, youth and marginalized groups on attitudes toward the state of the community, forms of inclusion and exclusion, power and leadership, governance and political culture, and conflict and people's participation in governance.

The baseline studies confirmed that the KALAHI-CIDSS project is working in some of the poorest, most deprived areas of the country and validated its focus on community empowerment given the lack of participation and voice of poor communities in determining their development priorities. Furthermore, the studies revealed significant variations in terms of poverty, social capital and governance among different municipalities highlighting the need for flexible and tailored implementation arrangements adopted by the project.

The surveys will be repeated during and after project implementation providing data to conduct a rigorous evaluation of the project's impact.

Sources: Chase and Holmemo (2005), Empowering Civic Participation in Governance (2006), and Parker and Holmemo (emails 5/24/2006).

Develop analysis matrix

Developing a matrix of analysis, such as the one below, to guide implementation of an evaluation is extremely useful. Such a matrix includes the issues to be covered, outcome and impact indicators, methodologies, and sources of data. It will help ensure that the evaluation will: yield the information required, illustrate how data will be captured and how evaluation components can be designed to cross-validate and triangulate information from various sources, and frame the analysis.

²⁴ For more information on gathering baseline data, see Woolcock et. al. presentation (2004).

Sample Matrix of Analysis

Issues	Indicators	Methodologies	Sources of data
Poverty levels of SF communities/districts	Percent of households in community or district below poverty line or consumption levels of extreme poor	Use preexisting poverty measures to analyze the distribution of social fund investments across districts according to poverty level	Oversampling national household income/ consumption survey (LSMS) in SF communities
Physical capital	Extent to which subprojects respond to community priorities	Propensity score matching to match random sample of social fund projects and corresponding households, stratified by type of project, with non-SF facilities and corresponding households based on geographic proximity and characteristics of facilities. Difference-in-differences estimates.	Qualitative beneficiary assessment
Human capital	Improved educational status, school attendance, years completed, dropout and retention rates (before and after and versus comparators)		Household and facilities surveys of both SF and non-SF schools and households. Need baseline or recall questions.
Sustainability of impact	Quality and quantity of infrastructure and services over time		Facilities surveys of both SF and non-SF facilities.
Cost effectiveness of subprojects	Average cost per new school, health post, water system versus comparator projects	Unit cost comparison	SF database and information from government ministries and municipal govts

Source: Adapted from Matrix of Analysis of Nicaragua Emergency Social Investment Fund Impact Evaluation 1998 included in Baker (2000).

Step 5: Obtain data

If data is not already available or does not adequately suit the requirements of the impact evaluation (because, for example, it doesn't contain the necessary outcome variables or coverage is insufficient), then it must be constructed. For an evaluation that will generate its own data, there are four general steps in development data: (i) designing and elaborating the collection instruments, (ii) sampling, (iii) fieldwork, and (iv) data management and access. Integrating quantitative and qualitative data collection instruments requires highly orchestrated coordination during each of these stages.

Designing the instruments

Data collection instruments vary widely from questionnaires that typically collect and record information in numeric form or pre-coded categories to descriptive text from interviews or focus group discussions. Annex 3 summarizes the main data collection instruments used in impact evaluation and their strengths and weaknesses.

Questionnaire development has an enormous influence on how respondents answer questions and, thus, on the validity of the information collected. If questions fail to measure the desired aspect of a problem, or if questions engender omission or ambiguous responses, the results of the evaluation will be skewed. The structure of the questionnaire itself can facilitate or complicate data gathering, and can have significant implications for the ability of evaluators to analyze the results. In developing new data collection instruments, care should be taken to ensure comparability among survey instruments in order to permit cross-referencing of data and comparisons across sectors, countries and regions.

Interview guides for the collection of qualitative data range from relatively simple, open and loosely-structured short thematic guides to those more highly structured, with direct closed questioning and occasionally multiple choice questionnaire forms. Experience suggests that designing only tightly structured, closed questions preempts “real” findings from emerging during fieldwork and can result in neglecting issues which should be probed further. Although closed questioning simplifies coding and quantification, open interviewing – letting the interviewee decide what to talk about and what criteria to use to judge project activities – is equally important. As with questionnaires, the way in which the question is asked directly influences the way it is answered. Responses must be solicited in such a way that they can be recorded in an orderly and intelligible manner.²⁵ Questionnaires and interview guides must be tailored to the particular group about which understanding is sought (e.g., community beneficiaries, government officials, service delivery personnel).

Draft versions of questionnaires and interview guides should be pilot tested in conditions that mirror as closely as possible the actual fieldwork. It is useful to have data entry programs ready at the time of the pilot to test their functionality, as well as a pilot sample across the different populations and geographical areas. Pilot testing of the data collection instruments will reveal whether they can reliably produce the required data and how to operationalize data collection procedures. Evaluators should test whether: (i) the questionnaire/questioning will provide the full range of required information; (ii) individual modules account for all activities and variables, and that there are no missing or redundant questions; (iii) individual questions are clearly worded, allowing no ambiguous responses or multiple interpretations; and (iv) all responses have been anticipated and coded.

Staff must be trained to collect the quantitative and qualitative data specific to the evaluation. To ensure that the information collected will meet evaluation objectives, qualified analysts and field and data managers should be involved in the development of the topics and questions, the pilot test, and the review of the data from the pilot test.

²⁵ Owen and Van Domelen (1998); Salmen (1999)

Designing and selecting the sample

The next step in designing the evaluation is establishing appropriate samples. Sampling procedures will help researchers to better control the precision of the evaluation estimates, and obtain the adequate amount of information to attain the required degree of precision. The sample design need not be complicated, but an experienced sampling specialist is best placed to determine appropriate sampling frames, sizes and selection strategies.

There is a tradeoff between sample size, which must be sufficiently large to ensure reasonable precision of the resulting impact estimates, and the cost of data collection. With quantitative methodologies, minimal sample size is typically determined through the use of statistical power calculations done during evaluation planning. These calculations relate the statistical precision of an estimate to the sample size and the variance of the outcome variable. Power calculations may reveal that measuring some outcomes will be particularly challenging and expensive (e.g., changes in school enrollment when enrollment is already near universal).

Sample selection involves numerous technical issues, a detailed discussion of which is beyond the scope of this toolkit. However, sample selection should follow some general rules:

- If the samples of different units of analysis are linked (e.g., facilities, communities and households), an adequate sample should ensure that the principal sampling unit is clearly defined and that the sample size is determined by the proper indicator;
- If the evaluation is interested in specific groups of the beneficiary population, there must be a sufficiently large population from which to draw a random sample, or some method must be employed to compensate for an inadequate population size (e.g., stratifying the population by subproject type and size, urban/rural, indigenous/rest of the population); and
- The element must be prepared for analysis (e.g., the relationship between the size of the sample and the population from which it is selected, non-responsive rates and other information used to inform sampling weights).

Rigorous sampling (although, again, it need not be complex or sophisticated) is equally important in qualitative analysis to ensure the quality and validity of data. The sample pool – in terms of numbers of subproject sites and beneficiaries per site - must be of adequate size and density to address principal themes and questions in terms of coverage, representativeness, diversity and depth.²⁶ Stratifying the beneficiary sample according to different social, economic, occupational, gender, age, and ethnic group criteria is important. A sampling expert, together with analysts, should guide the selection process

²⁶ Owen and Van Domelen (1998)

to ensure that the procedures established will result in the correct informants being selected.

The solicitation of views and opinions from a variety of different institutional actors - service providers, intermediary organizations, local governments, non-governmental organizations and other local institutions can illuminate the community-level institutional dimensions of social fund activities.

Sites and participants chosen for qualitative research should be linked with subprojects, households and communities surveyed in the quantitative evaluations in order to facilitate triangulation of findings.

Implementing the fieldwork

Qualitative and quantitative data collection fieldwork should be carefully managed. When planning fieldwork, it is important to:

- notify communities prior to fieldwork to maximize time spent in the field effectively;
- consider temporal events that may affect the success of fieldwork and validity of data collected (e.g., holidays, rainy season, school year calendar); and
- consider the timing of interviews - the duration and time of day should depend on what is convenient for participants.

The type of staff needed to collect data in the field will vary according to the objectives of the evaluation and methods used. Working with local staff with extensive experience in collecting data similar to that needed for the evaluation can greatly facilitate fieldwork operations. Knowledge of the geographic territory to be covered, the norms used in locating and approaching informants is very valuable. Quantitative and qualitative surveys will likely require similar teams of people, yet the skills of the interviewers required for each may be quite different (see Chapter 6 for team composition).

Data collection staff should be trained in: (i) the objectives of the study; (ii) the use of the instruments; and (iii) community norms and rules. Manuals should be used for training and as a reference during fieldwork. Managers should supervise closely the fieldwork and develop a communication system for field workers to easily channel questions and problems and report on progress. Analysts and sample designers should be available during the fieldwork for any problem with the questionnaires or the sample.

*Managing and accessing the data*²⁷

A good data management system should ensure the timeliness and quality of the evaluation data. Timeliness will depend on having as much integration as possible between data collection and processing so that errors can be verified and corrected prior to the conclusion of fieldwork. This can be accomplished by:

²⁷ Baker (2000)

- Applying consistency checks to test the internal validity of the data collected both during and after data gathering;
- Ensuring proper documentation is available to the analysts who will be using the data, including: (i) information needed to interpret the data (e.g., codebooks, data dictionaries, guides to constructed variables, and any needed translations); and (ii) information needed to conduct the analysis (e.g., evaluation objectives, methodologies, data collection instruments, sample selection, fieldwork description and guidelines for using the data).
- Making sure that data is clean and accessible before the analysis;
- Making available the evaluation database. Evaluations have a high public good and the data may be useful for other applications, for example to do additional follow-up work to assess long-term impacts; and
- An open data access policy that encourages good data documentation, but protects informant confidentiality.

Step 6: Analyze, report and disseminate data

Analysis

Many techniques are available for analyzing data and evaluators should choose tools appropriate to the research question and the audience one hopes to understand the findings. Cross-tabulations are simple, intuitive tools for analyzing information. Cross-tabulations, however, allow the consideration of only two variables at a time. Multivariate regression analysis can reveal relationships among many variables simultaneously, and highlight which relationships are statistically significant. Multivariate regression assumes causality in one direction. When directional causality is unknown, instrumental variables is the most common technique to tackle which variables cause which. This technique, however, is extremely difficult to implement in practice.²⁸

Qualitative methodologies can also be subject to a range of analytical approaches, from basic coding to advanced database management. Relatively simple analytical review can be as, and sometimes more, effective than highly complex analysis. Highly descriptive, anecdotal evidence does not readily translate into policy recommendations. To be useful to management, qualitative findings must demonstrate sufficient quantification, disaggregation, substantiation and validity.

Conducting as much analysis on-site during fieldwork and preparing preliminary site reports before leaving provides the opportunity to cross-check, verify, triangulate and feed back information while still in the community. Researchers thus need to be trained in field data analysis. Developing sound analytical frameworks to guide the fieldwork can greatly facilitate the presentation of results using charts, tables, ratings and graphs.²⁹

²⁸ Grootaert, Chase and Van Bastelaer (2003)

²⁹ Owen and Van Domelen (1998)

The analysis of evaluation data, qualitative and quantitative, typically takes longer than anticipated, particularly if: (i) data is not clean or accessible at the beginning of analysis; (ii) analysts are not experienced with evaluations; or (iii) there is an emphasis on capacity-building through collaborative work.

Reporting and dissemination

Several products, such as report synopses highlighting recommendations, presentations, and press releases, as well as the full report, should be produced from the analytical work - with each product tailored according to the audience for which it is prepared (e.g., government policy makers, program managers, donors, general public, journalists and academics). Outputs should include clear and practical recommendations, and be timed to influence key decisions regarding a project's future.

Dissemination of impact evaluation results is critical to their use. Reports should be planned as part of a broad dissemination strategy that includes presentations for various audiences, press releases, feedback to informants and making information available on the web. In addition to results, the actual data and careful documentation of methods of analysis should also be made public. Open access and discussion over data, methods and results can foster transparency and broad acceptance of the findings.

If an impact evaluation is to be used effectively, the social fund must be well-prepared to turn findings and recommendations into action. Such preparation includes: (i) planning how and when to determine, with stakeholders, changes to social fund/CDD operations; (ii) developing an action plan to implement agreed changes; (iii) monitoring adherence to the action plan; and (iv) monitoring the impact of changes made.

CONTRACTING AN IMPACT EVALUATION

Previous sections have addressed steps in preparing, designing and conducting impact evaluations of social fund and other CDD projects. This chapter discusses issues regarding institutional options for carrying out the evaluation and defining and procuring the required evaluation services. Contracting an impact evaluation is the responsibility of the project team.

Institutional options

Choosing an appropriate institution to carry out the evaluation is critical to ensure the impartiality and quality of results. Several factors should be considered: First, the decision to use a foreign firm or local entity depends in part on the desire to balance efficiency with local capacity building. A private firm may be more dependable and timely, yet the opportunity to build local evaluation capacity is largely lost. Where there is some local capacity to implement all or parts of an evaluation, international input and supervision can assist.³⁰ Second, there are also benefits and costs to using different types of entities (e.g., private consulting firms, university research institutes, non-governmental organizations and government agencies) to conduct an impact evaluation. Recommending one type of institution over another ultimately depends on capacity, expertise and experience accessible locally. While for some types of evaluations it may be appropriate to use an internal government agency, the specialized skills required to conduct an impact evaluation usually merit using a qualified external evaluator. Coordination arrangements that help ensure transfer of knowledge and analytical relevance, ownership and policy impact of the results reduce the costs of not using public agency. Third, institutional arrangements should seek to achieve the appropriate balance between involving users of the evaluations and maintaining the objectivity and legitimacy of results. An evaluation conducted independently of the institution implementing the project being evaluated conveys greater objectivity and enhances the credibility of its results. Yet in order for the evaluation process to inform policy, evaluation users (e.g., program managers, policy makers, non-governmental organizations and international development institutions) must be sufficiently involved.³¹

Evaluation team

Regardless of the type of institution contracted to conduct the evaluation, key skills need to be included on the evaluation team. Although the size of the team and some of the profiles will depend on the size and scope of the particular evaluation, the general composition and responsibilities of team members include:³²

- *Evaluation director/team leader:* Responsible for designing the overall evaluation, including refining the information needs and indicators, selecting the

³⁰ If program evaluations are to be conducted regularly, they must be based on local capacity. Thus, building such capacity as well as building ownership for the process and products of impact evaluation within developing countries is critical. (Rawlings, Sherburne-Benz and Van Domelen. 2004)

³¹ Baker (2000), Blomquist (2003)

³² Adapted from Baker (2000)

evaluation methodology, identifying the evaluation team, and ensuring appropriate integration of quantitative and qualitative data and methodologies. The evaluation director will coordinate and help carry out policy analysis.

- *Quantitative policy analyst*: an analyst (typically an economist) is needed for the quantitative methodology design and data analysis. He/she should be involved in writing the evaluation report.
- *Qualitative policy analyst*: This analyst (typically a sociologist or anthropologist) is responsible for qualitative methodology design and data analysis, and should be involved in writing the evaluation report.
- *Sampling expert*: Responsible for guiding the sample selection process. For the quantitative data, the sampling expert should be able to determine the appropriate sample sizes for the indicators established and their calculation power, select the sample, review the results of the actual sample versus the designed sample and incorporate the sampling weights for the analysis. For qualitative data, the sampling expert should guide the sample selection process in coordination with the analysts, ensuring that the procedures established guarantee that the correct informants are selected.
- *Survey designer*: This person or team is responsible for designing the data collection instruments and accompanying manuals and codebooks, and coordinating with the evaluation director to ensure that the data collection instruments will produce the data required for the analysis. This person/team should also be involved in pilot testing and refining the questionnaires and interviewing guides.
- *Fieldwork manager and staff*: The manager should be responsible for supervising the entire data collection effort, from planning the routes for the data collection to forming and scheduling the fieldwork teams (generally composed of supervisors and interviewers). Supervisors generally manage the fieldwork staff (usually interviewers, data entry operators and drivers) and are responsible for the quality of data collected in the field. Interviewers administer the questionnaires and field researchers conduct qualitative data collection and analysis.
- *Data managers and processors*: Design the quantitative data entry programs, enter the data, check its validity, provide the needed data documentation and produce basic results that can be verified by the data analysts.

Research teams should include members with a variety of institutional backgrounds as well as a mix of disciplines, gender and qualifications.

Social Fund Responsibilities

The social fund itself retains responsibility for key tasks. The input of the social fund manager with a designated internal evaluation manager or team is essential in: framing the information needs and indicators; drafting terms of reference; reviewing proposals; selecting and supervising the external contractor; providing overall guidance on the design and implementation of the evaluation; participating in progress reviews during fieldwork; critically reviewing draft reports; and developing an action plan for implementing agreed recommendations.

Terms of reference

Terms of reference for carrying out an impact evaluation should include the following information:

1. Background

Brief description of the project objectives, rationale, context, processes and the principles governing these as well as the current status and scope of the project.

2. Impact Evaluation Objectives

Outline of the areas of inquiry to be assessed. The evaluation objectives should be conveyed in as much detail as possible to avoid misunderstandings and unmet expectations.

3. Impact Evaluation Implementation

Description of the responsibilities of all involved entities, including, for example, the social fund, evaluation team, government oversight bodies, statistical agencies and coordination arrangements. A detailed list of what the firm contracted is expected to do which is likely to include the following:

- Design, test and finalize comprehensive quantitative and qualitative data collection instruments;
- Design adequate sampling methodologies and generate sample adequate to collect necessary quantitative or qualitative data;
- Train field staff in qualitative and quantitative data collection;
- Input and clean data collected from fieldwork; and
- Analyze impact evaluation data to achieve impact evaluation objectives.

4. Evaluation Design

Guidance on possible methodologies to use and integration of quantitative and qualitative techniques. Description of possible surveys and other available data sources. The guidance, however, should be sufficiently flexible to allow a firm to propose its own methodology.

5. Expected Outputs

Description of all deliverables and review arrangements. Typical outputs include:

- *Methodology report* describes the design of the evaluation (to be prepared after the firm has designed the quantitative and qualitative instruments), the indicators to be used, the matrix of analysis, and a sample of data collection instruments; and details the proposed tools and techniques for each component. Methodological issues should include: sample size and design, relative size of treatment and comparison samples, significance level, power calculations, and procedure for identifying treatment and comparison groups for qualitative evaluation design. It should present how quantitative and qualitative components of the evaluation are to be integrated.
- *Test phase report* contains a summary of the pilot study results, final versions of the data collection instruments, finalized focus group and other qualitative procedures and descriptions of all modifications made.
- *Study implementation or fieldwork report* covers the firm's experience collecting data. It provides feedback on aspects of data collection that were particularly effective and informative as well as problems encountered.
- *Draft and then final impact evaluation report* is a comprehensive review of impacts combining survey data and qualitative study outputs. It presents results, conclusions, recommendations and policy implications.

6. Duration and suggested timeframe for evaluation activities

7. Payment schedule

Procurement/contracting

Consultant services can be procured using a variety of methods, including direct solicitation of an institution, the compiling of a shortlist of potential entities, and open tendering.³³ The appropriate procurement method depends largely on the existence and depth of local capacity to carry out an impact evaluation.

Impact evaluation services are typically procured using lump sum (fixed price) contracts based on clearly specified outputs as detailed above. In writing the contract, it is important to leave flexibility to adjust to changes arising during implementation.

³³ For a full discussion of options, see Guidelines: Selection and Employment of Consultants by World Bank Borrowers (revised 2002).

CONCLUSION

The increasing use by development partners of results-based management combined with the learning-by-doing approach inherent in social fund/CDD projects argues for evaluating what interventions (or aspects of these interventions) work, what do not and why. Evaluating the impact of social fund and other CDD projects presents significant methodological challenges. But by following guidelines based on good practices and expert research, impact evaluation designs can adopt pragmatic and realistic approaches in the face of constraints. With careful planning and strong linkages to policymakers and other stakeholders, the increased use of impact evaluation will improve the performance of social fund/CDD projects and results on the ground.

ANNEX 1: QUANTITATIVE IMPACT EVALUATION METHODOLOGIES

After introducing the concepts of the counterfactual and bias from observable and unobservable characteristics in Chapter 1, this annex presents an overview of the principal quantitative techniques and estimation methods used in impact evaluation and their advantages and constraints.³⁴

Evaluation designs are determined by the methods used to identify the group of non-participants. They can be categorized as: *experimental* or *quasi-experimental*. These methods also incorporate estimation strategies. As no method is perfect, rigorous evaluations use a combination of techniques and triangulate results.

Experimental Designs

Experimental or randomized evaluation design involves defining a set of individuals (or communities or other unit of analysis) equally eligible and willing to participate in the program and randomly assigning them through a type of lottery into two groups: those who receive the intervention (treatment group) and those from whom the intervention is withheld (control group).

Experimental designs are widely considered the most robust of the evaluation methodologies as they offer the following *advantages*:

- ✓ The assignment process itself creates comparable treatment and control groups that are statistically equivalent. Provided sample sizes are large enough, randomization controls for *all* differences, observable and unobservable, between the treatment and control groups (i.e., bias is distributed evenly).
- ✓ Interpreting results is simple: program impact is the difference between the means of the samples of the treatment group and control group. Any difference can be causally linked to the program intervention.

In practice, however, there are several *constraints* that make experimental designs difficult to implement:

- ✗ Experimental design must be built into a program at its inception.
- ✗ The denial of benefits or services to otherwise eligible members of the population for the purposes of the study may be unethical and/or politically difficult.
- ✗ It can also be politically difficult to develop the criteria to determine who will receive intervention and who will not. This difficulty can lead to problems ensuring that the group assignment is strictly random.

³⁴ Rawlings et. al. (2004); Baker (2000); Prennushi, Rubio and Subbarao (2000); www.worldbank.org/poverty/impact

- ✦ The scope of the program may be so broad that there are no non-recipient communities, such as with a nationwide program or policy change.
- ✦ Individuals in treatment or control groups may change certain identifying characteristics during the experiment that could invalidate or contaminate results. For example, if people move in and out of a project area, they may move in and out of the treatment or control groups. Alternatively, people who were denied a program benefit may seek it through alternative sources, or those being offered a program may not take it up.
- ✦ Experimental designs require large sample sizes, and can be expensive and time consuming, particularly in the collection of new data.

Quasi-experimental Designs

Quasi-experimental designs apply statistical techniques to construct comparison groups that resemble treatment groups, at least in observed characteristics. These techniques include: (i) propensity score matching; (ii) pipeline matching; (iii) reflexive comparisons; and (iv) instrumental variables.

Matched Comparisons

Widely considered the next best alternative to experimental designs, matched comparison methods allow for the identification of the closest comparison group of a sample of non-beneficiaries to a sample of beneficiaries. Both groups should be matched on the basis of either observed characteristics or characteristics known or believed to affect program outcomes. Matching can be done on various units, such as households and communities, allowing for greater flexibility in evaluation design. Matched comparisons compare the treatment group with a comparison group that has not received the intervention, but is otherwise as similar as possible based on observed characteristics. Statistical controls (discussed below) can then be applied to address differences between treatment and comparison groups. Treatment and comparison groups can be matched either before or after the intervention.

Propensity Score Matching

The most widely used type of matching is propensity score matching in which the comparison group is matched to the treatment group using the predicted probability of participation given observed characteristics, the “propensity score”. The propensity score is calculated using observed characteristics of the treatment group that contribute to selection for intervention (e.g., the criteria a social fund uses to select its target communities). The treatment group’s score is then matched with the score of the comparison group that did not receive the intervention. The closer the propensity scores of the comparison group and the treatment group, the better is the match. To control for unobservable characteristics, the comparison group is often selected from the same region

as the treatment group (although this does not ensure that all selection issues are addressed), and is administered the same questionnaire as the treatment group by similarly trained interviewers.

The main *advantages* of using propensity score matching are that it:

- ✓ Is often relatively simple to carry out by combining results of national survey data with a separate beneficiary survey using the same questionnaire.
- ✓ Can be implemented with a single cross-section of data (though it is strengthened with the inclusion of baseline data).
- ✓ Generally costs less and has a shorter time frame than randomization or baseline and follow-up analysis.
- ✓ Circumvents the ethical and political issues around denying services and benefits to otherwise eligible recipients as under experimental designs.

The *disadvantages* of propensity score matching is that:

- ✗ Comparison group is matched to the beneficiary group based only on observable characteristics. It demands careful consideration of the extent to which unobserved differences between the two samples remain.
- ✗ Often requires an exhaustive questionnaire to accurately match treated and untreated populations based on their observable characteristics.
- ✗ Large household surveys with the required welfare information for both beneficiary and non-beneficiary groups do not always exist.
- ✗ It may not be possible to find a comparator community, even with the required data, without employing other matching methods.
- ✗ Demands careful consideration of the extent to which unobserved differences remain between two samples.

Box 12: Main steps propensity score matching

- i. Obtain a representative sample survey of participants and non-participants. A larger survey will better facilitate good matching. If the data for the two groups comes from different surveys, they must be comparable (*e.g.*, from the same questionnaire and survey period);
- ii. Identify the characteristics that make the communities likely to participate in the social fund or in each of the types of sub-projects;
- iii. Pool the two samples and generate a propensity function, which equates community characteristics with the likelihood that a community will submit a sub-project proposal and receive the intervention;
- iv. Use this propensity function to predict the probability of program participation for each community (*i.e.*, the probit coefficient). These are the propensity scores; and
- v. Match the beneficiary and non-beneficiary communities by their propensity scores. Some of the non-beneficiary sample may have to be excluded at the outset because they have a propensity score outside the range found for the beneficiary sample. For each participant in the treatment sample, find the observation in the non-participant sample that has the closest propensity score (the “nearest neighbor”), as measured by the absolute difference in scores.

Source: Ravallion (1999)

Pipeline Matching

This method matches beneficiary groups that have already received an intervention with those that have been selected to receive a future intervention as it is assumed these groups have similar characteristics. Indeed, the communities should be similar in unobservable characteristics (which make some communities more likely to apply for a subproject), and observable characteristics (which make communities more likely to be awarded a social fund project). This effectively corrects for the self-selection bias and targeting methods inherent in social funds.

The *advantages* of pipeline matching are that it:

- ✓ Accounts for unobservable variables.
- ✓ Generates baseline information on communities due to receive a social fund investment in the future.
- ✓ Circumvents the ethical and political issues around denying services and benefits to otherwise eligible recipients as under experimental designs.

There are, however, *disadvantages*:

- ✘ Researchers need to be certain that social fund eligibility criteria have not changed significantly over time.
- ✘ The project pipeline must contain a sufficient number of projects to allow a representative sample, as well as substitute projects in case those originally selected for the sample have already been completed when the evaluation begins.
- ✘ The possibility of characteristics that were not controlled (*i.e.*, residual bias) requires statistical adjustment, making the study more complex. This also requires greater expertise in the design, analysis and interpretation of the study to reduce the bias as much as possible.

Reflexive Comparisons

In a reflexive comparison, the counterfactual is constructed on the basis of the situation of program participants before the program. Participants are compared to themselves before and after the intervention and function as both the experimental and control group. The differences in outcomes are then estimated based only on changes across the time period in which the program took place.

Reflexive comparison has the following *advantages*:

- ✓ Particularly useful in evaluations of full-coverage interventions where there is no scope for a control group.
- ✓ Comparing the group to itself accounts for unobservable characteristics that would otherwise lead to bias. The observed changes between the baseline and the follow-up data cannot, therefore, be attributed to these factors and must be due to either the program intervention, or other factors that must be accounted for using econometric modeling.

Although seemingly simple, there are a number of *problems* with reflexive comparison that makes it the least rigorous type of impact evaluation.

- ✘ Requires baseline data.
- ✘ Difficult to predict where social funds will invest, making it difficult to gather baseline data.
- ✘ Requires extensive econometric modeling to account for such factors as changes that took place across the whole country which would bias results.
- ✘ The situation of program participants before and after the intervention may change due to many reasons independent of the program. Unless very carefully

done, reflexive comparisons may not be able to distinguish between the program and other external effects, thus compromising the reliability of results.

Instrumental Variables

Instrumental variables are used to control for bias due to unobservables. Instrumental variables are such that they determine program participation but do not affect outcomes attributable to the intervention, and thus can be used to establish causality when participation is correlated with outcomes of interest (e.g., social capital). First, the instrumental variables are used to predict project participation, and then the variation of the outcome indicator with the predicted values is analyzed. The difficulty in identifying good instrumental variables as well as the need for large survey samples constrain the use of this method.

Estimation techniques

Multivariate regression analysis

Multivariate regression analysis is used with matching techniques to control for possible observable characteristics that distinguish participants from non-participants. Selection bias is similar to an omitted variable bias in regression analysis. Thus, if it is possible to control for all reasons why outcomes might differ, then regression analysis can estimate the impact of the intervention, as well as the relative magnitude of different factors' influences on the overall welfare of the participants. This allows for information that can be used to develop complementary policies to address the same issue. The differences in the mean outcomes of the two groups, conditional on the set of variables that cause outcome and participation, constitute the effect of the intervention.

Difference-in-Differences

Difference-in-differences (or double difference) estimates can be applied in both experimental and quasi-experimental designs, and require baseline and follow-up data from the same sample treatment and control groups. The double difference method calculates mean differences between the treatment and control or comparison groups both before (first difference) and after (second difference) the intervention. To determine the first difference, a baseline survey is required before the intervention, and must cover both participant and non-participants in the program. Large national surveys can be helpful in obtaining this information.

After the intervention, a follow up survey is required. The baseline and follow-up surveys must be identical or, at least, highly comparable (e.g., same questionnaire or same indicators). Ideally, follow-up surveys should be of the same sampled observations as the baseline survey. If this is not possible, they should come from the same geographic clusters or strata to minimize bias. After gathering the information of the four groups, the analysis becomes simple. Evaluators simply calculate the mean difference between the before and after values for each treatment and comparison group, and then

calculate the difference between those two means differences. This will be the estimated impact of the program.

ANNEX 2: COST-EFFECTIVENESS ANALYSIS

Given its importance for measuring project efficiency, this annex provides more detail on conducting cost-effectiveness analysis.³⁵

Cost analysis calculates project costs compared to benefits and to alternative uses of the same resources. In doing so, it seeks to determine whether the program is more or less cost-effective than comparable interventions in delivering specific services (e.g., constructing a classroom) or, ideally, achieving a particular impact (e.g., raising enrollment or reducing infant mortality). Such information is critical to policy makers choosing among alternatives, particularly in resource-constrained environments.

Cost-effectiveness analysis of social fund projects has focused primarily on two levels: (i) unit costs of construction or rehabilitation of infrastructure at the subproject level; and (ii) project efficiency at a general level measured by overhead costs and process efficiency. Specifically, evaluations have addressed the following questions:

1. How do unit costs of social fund investments compare to investments in similar infrastructure constructed by other agencies?
2. What is general level of overhead or administrative costs associated with carrying out given investments and how do these compare to other agencies?
3. How efficient is the implementation process in terms of speed of execution?
4. What is the extent to which social funds and other social infrastructure programs leverage local resources in form of community contributions?

Also needed by policy makers, but not yet systematically monitored, is some concept of the unit costs of social funds in producing specific outcomes in comparison to other programs.

Despite the simplicity of its logic, cost analysis is particularly challenging to conduct – due primarily to difficulties ensuring programs are comparable and lack of data and standard accounting norms. Despite these challenges, however, cost analysis remains an important component of a robust impact evaluation.

Methodology

Define unit of analysis

As with impact evaluation methodologies, an important first step in cost analysis is to define the unit of analysis. To carry out relevant comparisons, a standard definition of investment is needed.

Agencies, however, define investments differently. For example, one agency's standard primary school construction program may include desks, blackboards, sanitary facilities; another's may include outdoor play spaces and retaining walls; while still another's may only

³⁵ Rawlings, Sherburne-Benz and Van Domelen (2002), WBER (2002), Blomquist (2003), Baker (2000).

cover basic construction. In addition, even when investments are uniformly defined, agencies may differ on the quality of inputs used.

Identify relevant comparators

There are many delivery mechanisms that countries use to finance small-scale infrastructure, including social funds, line ministries, local governments, NGOs and the private sector. These programs often exhibit differences in technical designs, procurement systems, quality of materials and approaches to community participation.

Again, finding valid comparator programs can prove extremely difficult, particularly when the social fund undertakes most of the country's community level investments in specific sectors. Even when an organization executing similar investments is found, it may have characteristics that complicate comparison to the social fund, such as its scope of action (working on a local, regional or national basis) which influence directly unit and overhead costs.

Calculate costs

Once the investments and delivery mechanisms are defined, cost-effectiveness analysis calculates all costs associated with delivering the investments. These costs typically include direct investment expenses and the administrative expenses incurred in running the program.

Cost analysis theory holds that total investment cost should also include the full value of community contributions, including the cost of donated materials and the opportunity cost of participating in the project.

Estimate benefits

Ideally, benefits are taken from the results of the impact evaluation. Because of the widespread lack of such data, costs are typically divided across a unit of benefit, e.g., cost per meter of construction or cost per beneficiary.

Calculate cost-effectiveness ratios

In principle, calculations and comparisons are straightforward. If the ratio of the cost to appropriate unit of benefit is reasonable compared to alternatives, the program is considered cost-effective.

Difficulties in calculating and comparing unit costs arise from several factors:

- (i) There is no standard practice for classifying or valuing expenses, either among programs within the same country or across countries. For example, while some agencies might consider the costs of project preparation or supervision investment costs, others might count them as administrative expenses. Resources

spent on community outreach and capacity building activities might be considered investments in long-term capacity or charged as overhead costs.

- (ii) Unit costs vary considerably across agencies, sectors and countries. These variations arise from significant differences such as location, materials used, intensity of use (i.e., number of beneficiaries), complexity, and number of subcomponents per investment.
- (iii) Reliable information on costs is often not available. Many potential comparator programs do not report administrative expenses.
- (iv) There are no standard criteria for accounting for community contributions (e.g., donated labor and materials), and few programs account for the full value of these inputs. Estimating such costs may not be easy where markets are shallow and dispersed, such as in rural areas. In practice, because of these obstacles and the fact that cost accounting is not precisely done for assessment purposes, community contributions are generally not valued and accounted for by agencies.
- (v) Some implementing agencies face national value-added taxes while others may enjoy exemptions that artificially lower their unit costs.

Given these difficulties, evaluations have tried various approaches. One approach compares estimated costs of standard infrastructure designs used by different agencies. This helps control for some of the variance among types of investments in some sectors, but it may only apply to a small range of investments carried out by the program. Alternatively, another approach would be to isolate a smaller set of investments that appear similar in scope and scale and carry out a case study of these investments, including site visits to assess quality, functionality and sustainability issues.

Difficulties in interpreting costs

Even when researchers can address comparability and data problems, their results may be prone to misinterpretation. Lower unit costs do not necessarily translate into more efficient investments, particularly when builders use sub-standard designs and materials that reduce the expected lifetime of an investment. Higher unit costs may be “better” if they result in longer amortization periods and better quality of services provided by infrastructure itself.

Other efficiency measures

In addition to unit costs, other measures of efficiency include several at the program level: (i) overhead expenses; (ii) speed of execution; and (iii) amounts leveraged of community resources – factors contributing to cost-effectiveness.

Overhead costs

The cost of an investment project consists of direct investment expenses plus administrative expenses incurred in running the program. These administrative, or overhead, expenses – such as personnel, consulting and other services, equipment, transportation, rent, utilities, supplies and communications – can be a significant factor in overall investment efficiency. As noted above, comparing these expenses across programs is difficult as programs use different terminologies and systems for classifying such expenses.

Speed of execution

The speed with which investments are executed directly affects a program's administrative costs and is another indicator of efficiency.

Community resources

Leveraging local contributions to investment costs can promote cost-effective investments. These contributions are typically cash or in-kind contributions from beneficiaries (usually donated labor and materials) or co-financing by local governments. Local counterpart contributions vary considerably, and there is no consistent policy on community cost sharing.

ANNEX 3: FRAMING THE OBJECTIVES

As discussed in Chapter 2, one of the most important and difficult steps in designing an impact evaluation is establishing evaluation objectives. This annex seeks to provide a framework for categorizing broad areas of inquiry relevant to most social funds and other CDD projects. The questions posed are not meant to be exhaustive, but rather suggestive of areas that future evaluations might explore.

The proposed framework divides impacts into three broad categories: (i) socio-economic impacts of subprojects on communities and households; (ii) impacts on social capital and the dynamics of beneficiary communities; and (iii) impacts on existing institutions and public sector management. The unevenness in the level of detail in the following three broad areas reflects the relative lack of experience with and empirical data on community dynamics and public sector management impacts, in particular.

Table 3: Indicative Areas of Impacts

<i>Socio-economic impacts</i>	<i>Social capital and community dynamics</i>	<i>Public sector institutions and management</i>
Living standards/primary welfare	Social capital	Decentralization
Infrastructure and services	Collective action	Sectoral coordination
Poverty targeting	Participation	Institutional autonomy
Cost-efficiency	Preference targeting	Budgetary effects

Although presented as separate categories, these three broad areas of impacts as well as the subcategories within them are inter-related and casual relationships are not always distinct. For example, subproject sustainability depends on community ownership as well as effective coordination with line ministries, and may in turn influence the accumulation of social capital. Social capital is viewed as important for well-being, and greater well-being is accompanied by increased social and civic involvement. The success of decentralization depends in large part on local capacity, and community demand for services may increase government and local capacity to meet the increased demand.

Some topics – the most notable being the nature of the participatory process that is central to realizing social fund and other CDD project objectives - appear both as areas of inquiry as well as factors affecting one or more types of impacts. Beneficiary participation is expected to ensure that benefits are better targeted, more relevant to beneficiary needs and distributed more equitably; community infrastructure is better built and maintained; government is more responsive; and citizens are more involved, informed and engaged in self-initiated development activities. Thus, the nature and extent of beneficiary participation is included as a factor affecting almost every type of impact as well as an impact to be evaluated in and of itself.

The importance of understanding the socio-economic and community impacts of social funds is widely recognized. But the success of a social fund as a component of a national poverty

reduction strategy also depends on its impact on public sector management more broadly. Understanding the role of social funds within the government framework is increasingly important as social funds become more permanent institutions. The tradeoffs between the immediate need for efficient and effective service delivery to the poor and longer-term objectives of building capacity in permanent government structures must be explicitly recognized and evaluated.³⁶

The evaluation methodologies discussed in this toolkit are most applicable to socio-economic impacts. Increasingly, they are being applied and adjusted to measure community dynamics and social capital impacts. Evaluating the impacts of social funds and other CDD projects on public sector management issues, however, will require different and evolving instruments and methodologies that are not covered here.³⁷

Note: all of the questions stated below should be made against the relevant comparator. For a particular question, the evaluation should specify whether the counterfactual is: (i) a comparison community receiving no intervention; or (ii) a community with a similar intervention implemented through a different entity (e.g., central government ministry, local authority, NGO, donor organization). Both questions are valid, but as discussed in Chapter 2, they entail different methodological approaches.

A. Socio-economic impacts

(i) Living standards/primary welfare

Key question: How do social funds affect living standards? What is the long-term sustainability of welfare impacts?

Study questions:

- Do physical enhancements improve utilization rates (in comparison to doing nothing)?
- Does beneficiary participation increase utilization?
- Do facility access and utilization improve education and health impacts?
- Do economic subprojects increase income?

Factors affecting living standards impacts:

Quality of infrastructure and availability of complementary inputs

Location of infrastructure (rural/urban, conflict zone)

Duration of employment

Household poverty levels and parent education levels

Extent to which subprojects respond to community priorities

Quality of community participation

Stock of household/community social capital

³⁶ Bhatia (2003)

³⁷ Wassenich and Whiteside (2004), Bhatia (2003)

Possible Indicators:

Improved **educational** status:

- Enrollment rates
- Education gap³⁸
- Household education expenditures
- Drop out and retention rates
- School absenteeism
- Student/teacher ratios
- Attendance rates
- Age in first grade
- Years completed
- Age-for-grade
- Regularity in teacher attendance
- Mean test scores

Improved **health** status:

- Health facility use
- Health facility operating hours
- Infant and child mortality
- Increased breastfeeding
- Attended deliveries
- Average expenditures on health care
- Child vaccination rates
- Diarrhea incidence
- Incidence of disease
- Malnutrition
- Prenatal care
- Water accessibility and availability
- Access to sanitary facilities
- Sewerage connection rates

Improved **economic** status:

- Increased income
- Increased employment

(ii) Quality and sustainability of infrastructure and services

Key questions: What is the quality of social fund-financed infrastructure, social services and economic activities? Are social fund investments maintained?

Study questions:

- Are social fund projects of higher quality and better managed than similar projects provided by central governments?
- Does the quality of works and local ownership enhance sustainability?

Factors affecting quality and sustainability:

Project sponsor/intermediary	Recurrent cost coverage
Complexity of design and project size	Maintenance arrangements
Use of technical standards	Availability of complementary inputs
Procurement process	Alignment with community priorities
Contractor qualifications	Quality of participatory process
Quality of construction	Community participation and training
Quality of supervision	Beneficiary satisfaction

Possible Indicators:

- Physical condition of infrastructure built/refurbished

³⁸ The education gap is the different between students' actual and ideal age.

- Space per student/patient
- Staffing levels
- Facility access to safe water, sanitation, electricity
- Availability of furniture, equipment and supplies
- Maintenance schedules and levels
- Payment of maintenance fees/cost-recovery
- Connection rates (water and sewerage systems)
- Beneficiary satisfaction
- Number of trainees
- Indicators of benefits of training
- Stock of social capital

(iii) Poverty targeting

Key questions: Do social funds reach poor and vulnerable communities and households?

Study questions:

- Are public goods delivered through social funds better targeted to the poor (than funds disbursed directly by the municipality)?
- Does community participation improve the targeting of private benefits like welfare or relief?
- Does the community-demand mechanism make it more difficult to reach poor communities which are often least competitive in preparing proposals?
- Does elite capture result in benefits accruing disproportionately to wealthier households within a community?
- Are social fund investments subject to less political influence than investments in government programs?

Factors affecting targeting performance:

Poverty map	Transparency of community decision-making
Promotional strategies	Heterogeneity of communities
Target resource allocations	Social capital of community
Eligible investments	Education level of household
Type of subproject	Degree of national income inequity
Level of targeting	Age of social fund/degree of autonomy
Community contribution	Distance to social fund headquarters
Beneficiary group access to info	Decentralized field offices

Possible indicators:

- Poverty levels of communities/districts and households
- Types of beneficiaries
- Distribution of resources (e.g., by geographic, household levels/per capita income)

- Subproject allocation (number of subprojects per district)
- Links between electoral targets and social fund spending

(iv) Cost-efficiency

Key question: How do social funds' costs compare to those of other institutions (e.g. line ministries, local governments, NGOs) carrying out similar projects?

Study questions:

- Does standardization increase speed and efficiency (compared to not standardizing)?
- Do community contributions result in lower costs?
- Does community management of resources and contracting lower unit costs?
- Do social fund subprojects have significantly lower levels of corruption in comparison to community-level investments implemented by alternative delivery systems?

Factors affecting cost-efficiency:

Structure of social fund
 Percentage of beneficiary contribution
 New construction v. rehabilitation
 Procurement process
 Speed of execution
 Transparency of social fund operations
 Beneficiary participation

Possible indicators:

- Average cost per new school, health center, water system, etc.
- Unit costs (cost per square meter of construction, per km of road, etc.)
- Average cost per beneficiary by project type
- Average cost of employment generated
- Social fund institutional costs as share of social fund projects
- Average completion time

B. Community dynamics and social capital impacts

(i) Social capital

Key question: Does social fund engagement at the community level through participation and training enhance social capital and thereby increase local capacity for collective action?

Study questions:

- Do social fund participatory processes strengthen community capacity and social capital (compared to traditional top-down approaches or to other participatory processes)?
- Does social cohesion/capital increase participation and facilitate collective action?
- Are communities with more social capital more productive and effective in their use of resources?
- Does social capital improve project effectiveness?
- Does the quantity and quality of group activity enhance the capacity of individuals to improve their own lives?

Factors affecting capacity building and social capital:

Quality of participation

Inclusiveness/exclusiveness of associations

Extent of transparency and accountability within local decision-making process of participating communities

Legal/institutional/political framework

Timing of participation with respect to life of social fund

Possible indicators:

- Membership in local associations and networks
 - density of associations (% of households participating in local organizations)
 - incidence of household membership (avg. # of memberships per household)
 - inclusiveness (internal diversity of association's membership)
 - institutional functioning (how decisions are made)
 - scope of network
 - access to information
- Measures of solidarity, trust and adherence to norms
 - extent to which one trusts people overall
 - extent to which households receive or would receive assistance from members in case of emergencies
 - numbers of disputes
 - community cohesion
- Indicators of collective action
 - quality and frequency of collective decision-making and action
 - capacity to address problems
 - capacity for future community-initiated activities
 - ability to contact and coordinate with other local organizations
 - ability to demand services from local government institutions
 - increased demand for "corruption-free" services
 - reduced dependence of communities on social fund to implement projects
- Increased transfers to community level organizations

- Income or social status of household

(ii) Quality of participation

Key Question: How participatory is the social fund decision-making process?

Study questions:

- Does participation lead to the empowerment of marginalized groups? (Does it lessen exclusion; increase capacity for collective action; and/or reduce potential for elite capture more than when they do not participate?)
- Does it allow for voice of excluded groups?
- Does it provide sufficient information on program rules of the game?
- Is there coherence within broader participatory local planning frameworks?
- Does involvement of the community improve its capacity for collective action?
- Does the possibility of greater community involvement and ownership result in more sustainable facilities?

Factors affecting quality of participation:

Characteristics of external agents (e.g., NGOs, facilitators)

Community openness to influences

Understanding of local power structures

Potential for collusion

Operational demands of subproject

Possible Indicators:

- Number and type of people attending and not attending committee meetings
- Number/type of marginalized people attending meetings
- Intensity of participation by different types of participants
- Number and type of decisions made when and how often
- Quality of decision-making
- Access to information
- Community awareness of and satisfaction with committee decisions
- Willingness to participate in future meetings/activities

(iii) Alignment of investments with community priorities (i.e., preference targeting)

Key Question: Do investments reflect the expressed priorities of the targeted communities and beneficiaries?

Study questions:

- Is the participatory approach effective in aligning investments with expressed community needs (more than centralized planning approaches)?
- Do benefits accrue disproportionately to wealthier households within a community?

Factors affecting alignment of priorities:

Beneficiary participation in subproject cycle stages
Community facilitation approach and quality of facilitator
Inclusiveness of participatory process
Access to key project information

Possible indicators:

- Alignment of *ex post* project delivery with *ex ante* expressed preferences of beneficiaries
- Numbers and types of complaints

C. Public Sector Management and Institutional Impacts³⁹

Key Questions: How does the creation of a social fund impact public sector management and existing institutions? Have social funds improved countries' capacities to use their human, organizational and financial resources effectively?

(i) Decentralization

Questions:

- Does the design and implementation of social funds support or undermine process of decentralization (compared to similar locales where no social fund is present)?
- Do social funds crowd out the allocation of funds to local governments?

Issues:

Social fund integration in decentralization framework/strategy
Extent to which social fund procedures are consistent with and reinforce those of decentralized structures
Integration of local government in social fund processes
Strength of support for decentralization
Level of political interference

Possible Indicators:

- Impact of social fund on institutional strengthening of local government
- Evolving role of local government in service provision and in relation to social fund role
- Shifting of responsibilities from social fund to decentralized agency
- Changes in social fund size and activities where decentralized agencies have adequate capacity
- Changes in local level decision-making
- Changes in intergovernmental fiscal framework

³⁹ Bhatia (2003), Parker and Serrano (2000)

(ii) Sectoral Coordination

Questions:

- Are social funds adequately coordinated with sector policies and investment priorities?
- Do social funds undermine sectoral coordination and weaken line ministries?

Issues:

National public sector reform strategy

Nature and extent of social fund linkages with line ministries

Explicit institutional development component

Possible Indicators:

- Adoption by line ministries of social fund innovations, operating procedures
- Capacity of line ministries to deliver social fund-type services
- Maintenance of social fund infrastructure
- Recurrent cost coverage
- Provision of complementary inputs in social fund infrastructure (e.g., staffing, equipment, supplies)

(iii) Social fund autonomy

Questions:

- How does the creation of an autonomous social fund affect the reform of public sector institutions?
- Do social funds promote wider adoption of values, behaviors, methods and processes and skills in national and local institutions?
- Do social funds distort public sector incentives?
- Does autonomy make the social fund more transparent, responsive, efficient and effective?

Issues:

Nature and extent of structural autonomy

Nature and extent of operational autonomy

Life span of social fund

Social fund salary structure compared to central government

Capacities of existing government agencies

Constituency for reform in central government/line ministries

Possible indicators:

- Effectiveness of social fund
- Impact of social fund provided technical assistance and training
- Adoption of innovative social fund procedures (e.g., targeting, technical standards, contracting, disbursement) by line ministries/agencies

- Involvement of line ministries/agencies in subproject planning and/or implementation
- Change in demand for public sector reform
- Effectiveness of other agencies to compete with social funds
- Government capacity to meet increased community demand
- Net gains/losses of movement of civil servants to social fund
- Quality of private sector participation

(iv) Budgetary Effects

Questions:

- What is the effect of social funds on national budgetary procedures?
- What is the aggregate impact of social fund activities on the overall balance and efficiency of public expenditure at national or local levels, within and among sectors, and between capital and recurrent expenditures?

Issues:

Transparency and accountability of social fund budgetary process
 Size of social fund budget relative to sectoral budgets and total budget
 Fungibility of funds/crowding out
 Incentives to repair vs. replace infrastructure

Possible indicators:

- Integration of social fund budgetary and audit procedures with central government
- Reduction in line ministry budgets in areas of overlap with social fund
- Shifting by communities from private to social fund provision of services
- Evolution of donor support for social activities implemented by government
- Aggregate amount of beneficiary cost-sharing
- Aggregate recurrent costs generated by social fund
- Change in capital vs. recurrent cost ratio

ANNEX 4: MAIN DATA COLLECTION INSTRUMENTS FOR IMPACT EVALUATION⁴⁰

<i>Technique</i>	<i>Definition and use</i>	<i>Strengths</i>	<i>Weaknesses</i>
Case studies	Collecting information that results in a story that can be descriptive or explanatory and can serve to answer the questions of how and why	<ul style="list-style-type: none"> - Can deal with a full variety of evidence from documents, interviews, observation - Can add explanatory power when focus is on institutions, processes, programs, decisions and events 	<ul style="list-style-type: none"> - Good case studies are difficult to do - Require specialized research and writing skills to be rigorous - Findings not generalizable to population - Time consuming - Difficult to replicate
Focus groups	Holding focused discussions with members of target population who are familiar with pertinent issues before writing a set of structured questions. The purpose is to compare the beneficiaries' perspectives with abstract concepts in the evaluation's objectives.	<ul style="list-style-type: none"> - Similar advantages to interviews (below) - Particularly useful where participant interaction is desired - A useful way of identifying hierarchical influences 	<ul style="list-style-type: none"> - Can be expensive and time consuming - Must be sensitive to mixing of hierarchical levels - Not generalizable
Interviews	The interviewer asks questions of one or more persons and records the respondents' answers. Interviews may be formal or informal, face-to-face or by telephone, or closed- or open-ended.	<ul style="list-style-type: none"> - People and institutions can explain their experiences in their own words and setting - Flexible to allow the interviewer to pursue unanticipated lines of inquiry and to probe into issues in depth - Particularly useful where language difficulties are anticipated - Greater likelihood of getting input from senior officials 	
Observation	Observing and recording situation in a log or diary. This includes who is involved; what happens; when, where, and how events occur. Observation can be direct (observer watches and records), or participatory (the observer becomes part of the setting for a time period of time).	<ul style="list-style-type: none"> - Provides descriptive information on context and observed changes 	<ul style="list-style-type: none"> - Quality and usefulness of data highly dependent on the observer's observational and writing skills - Findings can be open to interpretation - Does not easily apply within a short time frame to process change
Questionnaires	Developing a set of survey questions whose answers can be coded consistently.	<ul style="list-style-type: none"> - Can reach a wide sample, simultaneously - Allow respondents time to think before they 	<ul style="list-style-type: none"> - The quality of responses highly dependent on the clarity of questions

⁴⁰ Baker (2000) as adapted from Taschereau

		<p>answer</p> <ul style="list-style-type: none"> - Can be answered anonymously - Impose uniformity by asking all respondents the same things - Make data compilation and comparison easier 	<ul style="list-style-type: none"> - Sometimes difficult to persuade people to complete and return questionnaire - Can involve forcing institutional activities and people's experiences into predetermined categories
Written document analysis	Reviewing documents such as records, administrative databases, training materials, and correspondence.	<ul style="list-style-type: none"> - Can identify issues to investigate further and provide evidence of action, change, and impact to support respondents' perceptions - Can be inexpensive 	<ul style="list-style-type: none"> - Can be time consuming

REFERENCES

- Baker, Judy. 2000. *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Directions in Development. Washington, D.C.: World Bank.
- Bhatia, Mukhmeet. 2005. "Social Funds: A Review of Public Sector Management and Institutional Issues." SP Discussion Paper No. 0508. World Bank.
- Blomquist, John. September 2003 (draft). "Impact Evaluation of Social Programs: A Policy Perspective."
- Chase, Robert. 2006. "Thailand Social Capital Evaluation: A Mixed Methods Assessment of the Social Investment Fund's Impact on Village Social Capital." EASES. World Bank.
- Chase, Robert and Camilla Holmemo. 2005. "Community Driven Development and Social Capital: Designing a Baseline survey in the Philippines." SDD. World Bank.
- DECRG. (2006). "AAA Proposal for the Evaluation of NRSP: A Concept Note".
- Empowering Civic Participation in Governance. 2006. "Qualitative Baseline Survey for the Impact Evaluation of the KALAHI-CIDSS."
- Grootaert, Christian, Rob Chase and Thierry Van Bastelaer. 2003. "SOCAT Instruction Manual #3: Analysis." Social Development Department. World Bank.
- Grootaert, Christian and Thierry Van Bastelaer. 2002. *Understanding and Measuring Social Capital: A Multidisciplinary Tool for Practitioners*. Directions in Development. Washington, D.C.: World Bank.
- Jorgensen, Steen. 2005. "Framing the Logic of CDD operations: Towards a project theory for CDD design and evaluation." (draft)
- Kumar, Nalini. 2003. "Community-Driven Development: Lessons from the Sahel." Operations Evaluation Department.
- Kusek, Jody Zall and Ray C. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, D.C.: World Bank.
- Mansuri, Ghazala and Vijayendra Rao. 2003 "Evaluating CDD: A Review of the Evidence."
- Olken, Benjamin. 2004. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." NBER.
- OED. 2002. *Social Funds: Assessing Effectiveness*. Washington, D.C.: World Bank.

- OED. 2004. *Monitoring and Evaluation: Some Tools, Methods and Approaches*. Washington, D.C.: World Bank.
- OED. 2005. *The Effectiveness of World Bank Support for Community Development: An OED Evaluation*. Washington, D.C.: World Bank
- Owen, Daniel and Julie Van Domelen. 1998. "Getting an Earful: A Review of Beneficiary Assessments of Social Funds." Social Protection Discussion Paper 9816. Washington, D.C.: World Bank.
- Parker, Andrew and Rodrigo Serrano. 2000. "Promoting Good Governance through Social Funds and Decentralization." Social Protection Discussion Paper 0022. Washington, D.C.: World Bank.
- Prennushi, Giovanna, Gloria Rubio and Kalanidhi Subbarao. 2000. "Monitoring and Evaluation" in *Sourcebook for Poverty Reduction Strategies*. Washington, D.C.: World Bank.
- Rao, Vijayendra and Ana Maria Ibanez. 2001. "The Social Impact of Social Funds in Jamaica: A Mixed-Methods Analysis of Participation, Targeting and Collective Action in Community Driven Development."
- Rao, Vijayendra and Michael Woolcock. 2003. "Integrating Qualitative and Quantitative Approaches in Program Evaluation," in Francois Bourguignon and Luiz Pereira Da Silva (editors) *Evaluating the Poverty and Distributional Impact of Economic Policies: A General Introduction*. World Bank.
- Ravallion, Martin. 1999. "The Mystery of the Vanishing Benefits: Ms. Speedy Analyst's Introduction to Evaluation." Policy Research Working Paper 2153. World Bank.
- Rawlings, Laura, Lynn Sherburne-Benz and Julie Van Domelen. 2004. *Evaluating Social Funds: A Cross-Country Analysis of Community Investments*. Washington, D.C.: World Bank.
- Salmen, Lawrence F. 1999. "Beneficiary Assessment Manual for Social Funds." Social Protection Discussion Paper 9930. Washington, D.C.: World Bank.
- Valadez, Joseph and Michael Bamberger. 1994. *Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers and Researchers*. Washington, D.C.: World Bank
- Walker, Ian, Rafael del Cid, Fidel Ordonez and Florencia Rodriguez. 1999. "Expost Evaluation of the Honduran SIF (FHIS 2)" ESA Consultants, Honduras for World Bank.

Wassenich, Paul and Katherine Whiteside. 2004. "CDD Impact Assessments Study: Optimizing Evaluation Design Under Constraints." World Bank

Weissman, Julianna. 2001. "Operating Instructions Included: A Review of Social Investment Fund Operations Manuals." Social Protection Discussion Paper 0110. Washington, D.C.: World Bank.

White, Howard. 2006. "Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank." World Bank.

Wong, Susan. 2003. "Indonesia Kecamatan Development Program: Building a Monitoring and Evaluation System for a Large-scale Community-Driven Development Program." ESD EAP Discussion Paper. World Bank.

World Bank. 2002. "Guidelines: Selection and Employment of Consultants by World Bank Borrowers", Washington, D.C.: World Bank.

World Bank. 2000. "Nicaragua Ex-Post Impact Evaluation of the Emergency Social Investment Fund (FISE)", Report No. 20400-NI. World Bank.

World Bank. 2002. "Impact Evaluation of Social Funds" in The World Bank Economic Review Volume 16 2002 No. 2.

World Bank impact evaluation thematic group website: www.worldbank.org/poverty/impact

World Bank CDD evaluations project website: www.worldbank.org/cdd click on evaluations

Presentations:

Bamberger, Michael. Designing quality impact evaluations under budget, time and data constraints. July 28, 2005. BBL co-sponsored by OED and the Poverty Analysis, M&E Thematic Group.

Bamberger, Michael. Influential Evaluations. Evaluations that improved the performance and impacts of development programs. April 26, 2005. BBL co-sponsored by OED and the Poverty Analysis, M&E Thematic Group.

Chase, Rob. CDD Impact Evaluations: Practical Approaches to a Tough Analytic Problem. 2005. CDD and Social Capital Anchor.

Goldstein, Markus (2005) presentation to PADI workshop "An Introduction to Impact Evaluation." PRMPR.

Legovini, Arianna. International Experience in Impact Evaluation. Impact Evaluation Workshop. August 31, 2005.

Ravillion, Martin. DEC. Evaluating Anti-Poverty Programs. Part 1: Concepts and Methods.

Woolcock, Michael, Katherine Whiteside and Xavier de Souza Briggs. Assessing Development Projects: Integrating the Science of Evaluation with the Art of Practice. Poverty Impact Analysis, Monitoring and Evaluation Thematic Group. December 21, 2004.

Social Protection Discussion Paper Series Titles

<u>No.</u>	<u>Title</u>
0610	Risk and Vulnerability Considerations in Poverty Analysis: Recent Advances and Future Directions by Carlo Cafiero and Renos Vakis, October 2006
0609	Comparing Individual Retirement Accounts in Asia: Singapore, Thailand, Hong Kong and PRC by Yasue Pai, September 2006 (online only)
0608	Pension System Reforms by Anita M. Schwarz, September 2006 (online only)
0607	Youth Labor Market in Burkina Faso: Recent Trends by Daniel Parent, July 2006
0606	Youth in the Labor Market and the Transition from School to Work in Tanzania by Florence Kondylis and Marco Manacorda, July 2006
0605	Redistributing Income to the Poor and the Rich: Public Transfers in Latin America and the Caribbean by Kathy Lindert, Emmanuel Skoufias and Joseph Shapiro, August 2006 (online only)
0604	Uninsured Risk and Asset Protection: Can Conditional Cash Transfer Programs Serve as Safety Nets? by Alain de Janvry, Elisabeth Sadoulet, Pantelis Solomon and Renos Vakis, June 2006
0603	Examining Conditional Cash Transfer Programs: A Role for Increased Social Inclusion? by Bénédicte de la Brière and Laura B. Rawlings, June 2006 (online only)
0602	Civil-service Pension Schemes Around the World by Robert Palacios and Edward Whitehouse, May 2006 (online only)
0601	Social Pensions Part I: Their Role in the Overall Pension System by Robert Palacios and Oleksiy Sluchynsky, May 2006 (online only)
0543	Complementing Natural Disasters Management: The Role of Social Protection by Renos Vakis, February 2006

- 0542 Brazil's Bolsa Escola Program: The Role of Local Governance in Decentralized Implementation
by Alain de Janvry, Frederico Finan, Elisabeth Sadoulet, Donald Nelson, Kathy Lindert, Bénédicte de la Brière and Peter Lanjouw, December 2005 (online only)
- 0541 Japan's Pension Reform
by Junichi Sakamoto, December 2005 (online only)
- 0540 Demographic Alternatives for Aging Industrial Countries: Increased Total Fertility Rate, Labor Force Participation, or Immigration
by Robert Holzmann, December 2005 (online only)
- 0539 Disability, Poverty and Schooling in Developing Countries: Results from 11 Household Surveys
by Deon Filmer, November 2005 (online only)
- 0538 Food Aid and Food Security in the Short- and Long Run: Country Experience from Asia and sub-Saharan Africa
by Carlo del Ninno, Paul A. Dorosh and Kalanidhi Subbarao, November 2005 (online only)
- 0537 The New Pensions in Kazakhstan: Challenges in Making the Transition
by Richard P. Hinz, Asta Zviniene and Anna-Marie Vilamovska
September 2005 (online only)
- 0536 Household's Vulnerability to Shocks in Zambia
by Carlo del Ninno and Alessandra Marini, September 2005 (online only)
- 0535 Linking Community Empowerment, Decentralized Governance, and Public Service Provision through a Local Development Framework
by Louis Helling, Rodrigo Serrano, David Warren, September 2005 (online only)
- 0534 Youth Employment in the MENA Region: A Situational Assessment
by Nader Kabbani and Ekta Kothari, September 2005
- 0533 Measuring Risk Perceptions: Why and How
by Joachim De Weerd, July 2005
- 0532 Implementing Means-Tested Welfare Systems in the United States
by Kathy Lindert, June 2005 (online only)
- 0531 La Focalización En El Programa De Desarrollo Humano Oportunidades De Mexico

- by Mónica Orozco y Cecilia Hubert, June 2005 (online only)
- 0530 Sistema de Identificación de la Población Objetivo: SIPO en Costa Rica
by Roxana M. Víquez, June 2005 (online only)
- 0529 Targeting Social Spending To the Poor with Proxy–Means Testing: Colombia's SISBEN System
by Tarsicio Castañeda, June 2005 (online only)
- 0528 Focalización De Programas En Chile: El Sistema CAS
by Osvaldo Larrañaga, June 2005 (online only)
- 0527 Reforming Brazil's Cadastro Único to Improve the Targeting of the Bolsa Família Program
by Bénédicte de la Brière and Kathy Lindert , June 2005 (online only)
- 0526 Designing and Implementing Household Targeting Systems: Lessons from Latin American and The United States
by Tarsicio Castañeda and Kathy Lindert, with Bénédicte de la Brière, Luisa Fernandez, Celia Hubert, Osvaldo Larrañaga, Mónica Orozco and Roxana Viquez, June 2005 (online only)
Also available in Portuguese.
- 0525 Unequal Prospects: Disparities in the Quantity and Quality of Labour Supply in sub-Saharan Africa
by John Sender, Christopher Cramer and Carlos Oya, June 2005
- 0524 Pension Supervision: Understanding International Practice and Country Context
by Richard P. Hinz and Anca Mataoanu, May 2005
- 0523 Social Security Coverage in Latin America
by Rafael Rofman, May 2005
- 0522 A Technology White Paper on Improving the Efficiency of Social Safety Net Program Delivery in Low Income Countries: An Introduction to Available and Emerging Mobile Technologies
by Mike Gallaher, May 2005 (online only)
- 0521 Aging and Poverty in Africa and the Role of Social Pensions
by Nanak Kakwani and Kalanidhi Subbarao, May 2005 (online only)
- 0520 Delivery Mechanisms of Cash Transfer Programs to the Poor in Bangladesh
by Shaikh S. Ahmed, May 2005 (online only)

- 0519 Portability Regimes of Pension and Health Care Benefits for International Migrants: An Analysis of Issues and Good Practices
by Robert Holzmann, Johannes Koettl and Taras Chernetsky, May 2005
(online only)
- 0518 Who Has the Yam, and Who Has the Knife? Social Action Funds and Decentralization in Malawi, Tanzania, and Uganda
by N. Mungai Lenneiye, May 2005
- 0517 Household Vulnerability and Children's Activities: Information Needed from Household Surveys to Measure their Relationship
by Diane Steele, May 2005 (online only)
- 0516 The Effect of Child Labor on Mathematics and Language Achievement in Latin America
by Mario A. Sanchez, Peter F. Orazem, and Victoria Gunnarsson, May 2005
- 0515 The Inter-Generational Persistence of Child Labor
by Patrick M. Emerson and André Portela Souza, May 2005
- 0514 How Does Working as a Child affect Wage, Income and Poverty as an Adult?
by Nadeem Ilahi, Peter F. Orazem, and Guilherme Sedlacek, May 2005
- 0513 Dynamics of Child Labor: Labor Force Entry and Exit in Urban Brazil
by Suzanne Duryea, Jasper Hoek, David Lam and Deborah Levison, May 2005
- 0512 The Responses of Child Labor, School Enrollment, and Grade Repetition to the Loss of Parental Earnings in Brazil, 1982-1999
by Marcelo Côrtes Neri, Emily Gustafsson-Wright, Guilherme Sedlacek and Peter F. Orazem, May 2005
- 0511 Child Labor, Schooling, and Poverty in Latin America
by Guilherme Sedlacek, Suzanne Duryea, Nadeem Ilahi, and Masaru Sasaki, May 2005
- 0510 Changing Patterns of Child Labor around the World since 1950: The Roles of Income Growth, Parental Literacy and Agriculture
by Victoria Gunnarsson, Peter F. Orazem, and Guilherme Sedlacek, May 2005
- 0509 Disability and Social Safety Nets in Developing Countries
by Sophie Mitra, May 2005

- 0508 Social Funds: A Review of Public Sector Management and Institutional Issues
by Mukhmeet Bhatia, May 2005
- 0507 Pension Reform in El Salvador
by Rodrigo Acuña, April 2005
- 0506 How Changes in Benefits Entitlement Affect Job- Finding: Lessons from the Slovenian “Experiment”
by Jan van Ours and Milan Vodopivec, April 2005
- 0505 A Guide to Multisector Labor Market Models
by Gary S. Fields, April 2005
- 0504 Multinational Enterprises and Training Revisited: Do International Standards Matter?
by Niels-Hugo Blunch and Paula Castro, March 2005
- 0503 Community-Based Health Insurance & Social Protection Policy
by Steven Tabor, March 2005 (online only)
- 0502 Towards a Better Understanding of the Nature, Causes and Consequences of Youth Labor Market Disadvantage: Evidence for South-East Europe
by Alexandre Kolev and Catherine Saget, March 2005 (online only)
- 0501 A Lecture on the Political Economy of Targeted Safety Nets
by Lant Pritchett, January 2005
- 0420 Can Conditional Cash Transfer Programs Improve Social Risk Management? Lessons for Education and Child Labor Outcomes
by Elisabeth Sadoulet, Frederico Finan, Alain de Janvry and Renos Vakis, December 2004
- 0419 Measuring Welfare for Small but Vulnerable Groups. Poverty and Disability in Uganda
by Johannes G. Hoogeveen, September 2004
- 0418 Institutional Analysis Toolkit for Safety Net Interventions
by Inke Mathauer, August 2004
- 0417 Trade Union Participation in the PRSP Process
by Lawrence Egulu, August 2004
- 0416 A New Approach to Social Assistance: Latin America’s Experience with Conditional Cash Transfer Programs
by Laura B. Rawlings, August 2004

- 0415 Shocks and Coffee: Lessons from Nicaragua
by Renos Vakis, Diana Kruger and Andrew D. Mason, July 2004
- 0414 Costs of Projects for Orphans and other Vulnerable Children: Case studies in
Eritrea and Benin
by Menahem Prywes, Diane Coury, Gebremeskel Fesseha, Gilberte
Hounsounou, and Anne Kielland, July 2004
- 0413 Disability Employment Policy
by Daniel Mont, July 2004
- 0412 Child Labor, Education, and Children's Rights
by Gordon Betcherman, Jean Fares, Amy Luinstra, and Robert Prouty, July
2004
- 0411 Challenges and Opportunities of International Migration for the EU, Its
Member States, Neighboring Countries and Regions: A Policy Note
by Robert Holzmann and Rainer Münz, June 2004
- 0410 Evaluating Different Approaches to Estimating Vulnerability
by Ethan Ligon and Laura Schechter, June 2004
- 0409 Transferring Cash Benefits Through the Banking Sector in Colombia
by Maria Teresa Lafaurie V. and Claudia A. Velasquez Leiva, May 2004
- 0408 Mitigating Social Risks in Kyrgyz Republic
by Emil Tesliuc, April 2004
- 0407 Toward a Reformed and Coordinated Pension System in Europe: Rationale
and Potential Structure
by Robert Holzmann, April, 2004
- 0406 Boosting Productivity Via Innovation and Adoption of New Technologies:
Any Role for Labor Market Institutions?
by Stefano Scarpetta and Thierry Tressel, March 2004
- 0405 Mitigating the Social Impact of Privatization and Enterprise Restructuring
by David H. Fretwell, March 2004
- 0404 Risk and Vulnerability in Guatemala: A Quantitative and Qualitative
Assessment
by Emil D. Tesliuc and Kathy Lindert, March 2004
- 0403 Implicit Pension Debt: Issues, Measurement and Scope in International
Perspective
by Robert Holzmann, Robert Palacios and Asta Zviniene, March 2004

- 0402 Impacts of Active Labor Market Programs: New Evidence from Evaluations
with Particular Attention to Developing and Transition Countries
by Gordon Betcherman, Karina Olivas and Amit Dar, January 2004
- 0401 Consumption Insurance and Vulnerability to Poverty: A Synthesis of the
Evidence from Bangladesh, Ethiopia, Mali, Mexico and Russia
by Emmanuel Skoufias and Agnes R. Quisumbing, January 2004

**To view Social Protection Discussion papers published prior to 2004, please visit
www.worldbank.org/sp.**



Summary Findings

In response to the growing demand for evidence of how social fund (and other CDD) projects achieve significant and measurable improvement in peoples lives, this toolkit aims to increase the number, expand the topical coverage and improve the quality of social fund/CDD impact evaluations. Designed for task teams and managers of social fund and other CDD projects, as distinct from evaluation practitioners, the toolkit first defines impact evaluation and introduces related concepts and methodologies and then guides this audience through the process of planning, designing and implementing an impact evaluation.

HUMAN DEVELOPMENT NETWORK

About this series...

Social Protection Discussion Papers are published to communicate the results of The World Bank's work to the development community with the least possible delay. The typescript manuscript of this paper therefore has not been prepared in accordance with the procedures appropriate to formally edited texts. The findings, interpretations, and conclusions expressed herein are those of the author(s), and do not necessarily reflect the views of the International Bank for Reconstruction and Development / The World Bank and its affiliated organizations, or those of the Executive Directors of The World Bank or the governments they represent.

The World Bank does not guarantee the accuracy of the data included in this work. For free copies of this paper, please contact the Social Protection Advisory Service, The World Bank, 1818 H Street, N.W., Room G7-703, Washington, D.C. 20433-0001. Telephone: (202) 458-5267, Fax: (202) 614-0471, E-mail: socialprotection@worldbank.org or visit the Social Protection website at www.worldbank.org/sp