

Monitoring and evaluation

The objective of this chapter is to discuss the basic techniques for monitoring, evaluation and impact assessment relevant for social transfer programmes. In particular, the chapter addresses three questions:

What are they? Scientific techniques for measuring the implementation and outcomes of social transfer programmes.

When do we use them? They should be used from the beginning of the programme to assess the conceptualisation, design, implementation, successes and failings – both in order to improve delivery and to provide assessments of impact.

How do we use them? This chapter provides an eight-step framework for implementing an impact evaluation.

Introduction

Once social transfer schemes are designed and implemented, an ongoing process of monitoring and evaluation can improve delivery processes, document results, inform policymakers about the effectiveness of alternative approaches, and mobilise political support for programme sustainability and expansion. Monitoring is the process of identifying and tracking performance indicators and reviewing implementation over the life of the programme.¹ It can be defined as a continuing process of collecting and reviewing data on performance indicators in order to inform managers (and other stakeholders) about the progress and achievement of objectives of the programme.² Evaluation links causes to outcomes, assessing impacts after a programme (or distinct activities) have been completed.³ The evaluation of social transfers involves the objective and systematic assessment of the design, implementation and results of the programme.⁴ This chapter examines the major tools and

Box 15.1: Types of evaluation strategies

No single approach to evaluation is appropriate in every context. Some evaluations address impacts, while others are more concerned with process.

Budgets and time constraints vary across situations.

Seven major categories of evaluation strategies are:

- **Impact evaluations** aim to reveal which impacts were caused by the social transfer and which ones might have been caused by other factors. It is often regarded as the classic form of evaluation. This is a relatively complex and expensive form of evaluation, but it potentially produces high-quality and credible results.
- **Process implementation evaluations** are concerned with how the programme operates and focus on problems in service delivery, policy instruments, management practices and linkages amongst these factors. This evaluation method is similar to monitoring.
- **Rapid appraisal allows** for quick assessment and reporting. It can provide immediate feedback on the development of a programme, project or policy. This tends to be used for flagging problems during the implementation phase of the social transfer.
- **Meta-evaluation** establishes the criteria and procedures which can be used to systematically summarise findings from several existing evaluations. It is a technique for reviewing current evaluations in a rapid but integrated manner.

methodologies for monitoring and evaluating social transfer programmes and provides a framework for implementing an evaluation. The discussion aims to facilitate the role of monitoring, evaluation and impact assessment in supporting evidence-based policy design and implementation of social transfer programmes.

Monitoring and evaluation serve three major types of objectives:⁵

1. Strategic objectives: is the social transfer programme achieving its goals?

Governments and donor agencies have limited resources and competing priorities. Monitoring and evaluation can serve to ensure that programmes utilise funds in a justifiable and efficient manner in order to promote development and support the strategic objectives of social protection. Social transfer impact assessments compare outcomes for beneficiaries to the outcomes for non-beneficiaries.⁶

2. Operational objectives: how can programme managers improve implementation?

Monitoring (sometimes referred to as “formative evaluation”) tracks the progress and process of a social transfer programme’s implementation, mainly in order to check progress against a plan. Effective programme monitoring identifies and measures performance indicators in order to provide periodic feedback on the success of implementation as well as indications of problems that arise.⁷

3. Learning objectives: what can be learned from the social transfer programme?

Are there alternatives to the programme’s design and implementation processes? Does the programme illustrate best practices that provide a

- **Performance logic chain assessments** evaluate how the programme activities are organised and sequenced, and how the resources are utilised. The evaluation assesses how likely it is that the desired change in poverty level will be successfully achieved by social transfers.
- **Pre-implementation assessments** evaluate whether there is a coherent and clear implementation plan and whether the objectives and outcomes are clearly linked in order to ensure that the design of the programme includes all critical components.
- **Case studies** provide in-depth information about specific parts of the programme design or about particular beneficiary groups. Case

studies offer an in- depth analysis of specific examples, at the expense of providing a breadth of knowledge. They also tend to be more qualitative in nature.

This chapter focuses on impact evaluations and process implementation evaluations (monitoring). In addition, Box 5.1 provides an example of rapid appraisal in the context of public works. For more information about the other types of evaluation see Baker (2000) and Kusek and Rist (2004).

SOURCE: Kusek and Rist (2004), pages 121–126.

model for learning? What lessons can be drawn from the experience? Evaluations with learning objectives do not necessarily serve the specific programme managers or other stakeholders – they may aim to benefit the broader international community concerned with the design and implementation of social transfer programmes.

No single evaluation framework serves the needs of every programme – there are a variety of tools and methodologies from which to choose. Each evaluation’s design is unique, and depends on data availability, technical capacity of evaluators, time and budget constraints, as well as local political, economic, cultural and administrative factors. The most rigorous evaluations tend to be planned during the design phase of the programme (see chapter 3, section 5).⁸

In addition, “evaluations that will yield high-quality, credible, and generalisable results for policymakers will require strong financial and political support; early and careful planning; participation of stakeholders in the design of the objectives and approach of the study; adequate data; a suitable mix of methodologies, including both quantitative and qualitative techniques; the rigorous application of these techniques; and communication between team members throughout the process.”⁹

Evaluations are conducted within a political context where all the stakeholders have strong views concerning ethical issues, the choice of “correct” evaluation methods and how the findings should be disseminated. Faced with these and other limitations – including budget and time constraints, limited access to baseline data and comparison groups, lack of experienced or trained personnel to carry out the evaluation – evaluations may be difficult to conduct with robust designs and techniques. Often programme managers face a trade-off between available resources, and acceptable standards of evaluation practice and methodological rigour.¹⁰

Box 15.2: The potential and pitfalls of government-mandated evaluations

After the initial evaluations of Mexico's *Progresa* programme (now called *Oportunidades*), the government passed legislation in 2001 requiring annual evaluations for major social programmes. While the law does not directly provide incentives for quality evaluations, individuals in responsible state agencies often take a serious interest in the design and implementation. A group of public officials with experience and a strong interest in rigorous evaluation techniques designed and implemented the *Progresa* programme. They selected a randomised experimental methodology, providing a rigorous impact evaluation that was innovative in the region at the time. This design required treatment and control groups, involving the withholding of benefits from the control group for 18 months.

Progresa officials anticipated that legislators in Congress would react adversely, and they documented this design feature with intentional vagueness

until the evaluation was practically complete. Once Congress and the media understood the nature of the random experiment, a political controversy ensued, but it was too late to affect the evaluation process.

Even before the 2001 legislation required annual assessments, Mexico's Ministry of Agriculture commissioned the Food and Agricultural Organisation (FAO) of the United Nations to evaluate *Alianza para el Campo* and PROCAMPO (which provided cash payments to certain producers in the Mexican countryside) to assist in the design of an evaluation. FAO advisors worked with Ministry officials to design a quantitative evaluation of perceptions of impact. The evaluation design included questions for beneficiaries about their opinions regarding the impact of the programmes on productivity and income. The evaluations were conducted on an extensive scale – covering nearly all of the programmes in *Alianza* across each of the 32 states

Programmes or projects designed for monitoring and evaluation can answer the following questions:¹¹

- Does the programme reach the intended beneficiaries?
- How does the programme affect the beneficiaries? (The assessment should potentially evaluate nutrition, poverty, vulnerability to starvation, health, personal safety, ability to plan and cope with shocks, relations with others in the community, and gender inequalities.)
- Does the programme generate the desired outcomes?
- What is the impact on the rest of the population?
- Are there better ways to design the programme?
- Can the programme be managed more efficiently?
- Are the allocated resources being spent efficiently?
- Were the costs of setting up and running the programme justified?

For each of these questions both **qualitative data** (data expressed in narratives or words – using observations, interviews or focus groups) and **quantitative data** (data expressed in numbers using questionnaires or existing databases) can be useful in a variety of ways. While some researchers may differ on the respective merits of the two approaches, it is generally acknowledged that some combination of the two methods of data collection is important in any evaluation in particular when the programme is in a complex social environment, as is the case with any social transfer programme. There are four key reasons for using the mixed method approach: 1) it allows the evaluator a wide range of methods and tools to draw from, 2) it increases the validity of conclusions through “triangulation” by providing multiple

in Mexico. The data collection from tens of thousands of informants cost over US\$1 million per year – but the results proved fairly limited, particularly in consideration of the amount of resources spent. The evaluations could not answer some of the most important questions – for example, whether or not Alianza was responsible for a statistically significant impact on important social and economic indicators.

The Ministry of Agriculture's approach to compliance with the law mandating annual evaluations yielded undesirable outcomes in two ways – an excessive amount of data was collected in a costly manner, but without providing the essential information necessary for an effective evaluation. The bureaucratic interpretation led the Ministry to collect and process substantial amounts of data at a state level – leading to a costly report which failed to address some of the most critical issues. Over time, the Ministry expanded its interpretation of the

evaluation requirements and began to produce more useful assessments.

While the government-mandated evaluations have not always proven useful, academics and researchers have conducted several studies of PROCAMPO which support comparisons with *Progresa*. National household surveys conducted in 1994 and 1997 included a significant proportion of households benefiting from PROCAMPO. In addition, extensive surveys and studies of *Progresa* provide a wealth of information about the programme's impact. Along many dimensions the two programmes have demonstrated comparable effects. The impacts of the programmes appear similar in terms of improvements in household well-being, investment in productive activities, calorie intake, food consumption, and nutritional diversity.

SOURCE: Davis (2003).

independent findings that can be compared, 3) it allows for a deeper and richer analysis and understanding of the programme context, 4) it provides opportunities for reducing cost and time of data collection.¹²

There are many approaches, strategies and types of evaluation and no single method of evaluation is best suited to all purposes and all programmes. The most appropriate method of evaluation for a given programme depends on many factors, including: the type of answers needed; the nature of the programme being evaluated; the availability of data needed for the evaluation; how certain the policymaker needs to be about the data and the conclusions produced by the study; what level of resources is being devoted to the evaluation; and the time that is available for the study. Choosing an evaluation strategy will require trade-offs between time, cost and accuracy or confidence in the study.

Design issues in impact evaluation

Numerous methods and analytical techniques have been used in evaluation studies. Box 15.1 provides an overview of some of the major types of evaluation strategies. A comprehensive approach can include monitoring, process evaluation, cost-benefit analysis, and impact assessment.¹³

Combining complementary techniques spans the entire life of the social transfer programme and provides a more holistic evaluation. Impact evaluations assess whether the programme produced the intended effects on individuals and households, and determines if the effects can be directly attributed to the social transfer intervention. Impact evaluations can also

Box 15.3: The importance of a credible comparison group in impact assessment

The International Food Policy Research Institute (IFPRI) conducted an impact assessment of conditional cash transfers in Nicaragua from 2000 to 2002. Households benefiting from the cash transfers received an average of 3 500 Cordobas in the first year, and 3 800 in the second year. Annual household expenditure for these beneficiaries, however, increased only by 404 Cordobas over the two years – an insignificant impact, particularly given the cost of the programme. Can policy analysts conclude the programme was a failure?

No, the problem is not that simple. Simply measuring the change in a welfare measure over time for beneficiary households ignores other

factors affecting households. In this study, IFPRI also measured the change in household expenditure for a similar group of households that did not have the benefit of the conditional cash transfer programme. Their average expenditure fell significantly over the two years. It is likely that the average expenditure of the participating households would have fallen also in the absence of the programme. The significant expenditure stabilising impact of the conditional cash transfers is a mark of programme success.

This example demonstrates the importance of a “control group” in evaluating social transfer programmes. Comparing the living standards of households after the introduction of a social transfer

examine the range of consequence for beneficiaries – positive and negative, direct and indirect, intended and unintended. Monitoring assesses whether the programme’s implementation follows its plan and achieves the intended objectives. Process evaluation focuses on the implementation details of the programme. Cost-benefit analyses weigh monetary and non-monetary costs against quantifiable benefits – cost-effectiveness evaluations provide a similar exercise when benefits cannot be quantified.¹⁴

“Impact can be defined as the change, effect or benefit that results from the services or activities on a wider society than its direct users. It is often long-term, broad and sustainable and can include affecting policy decisions at government level.¹⁵ Impact evaluation “is the process of identifying the anticipated or actual impacts of a development intervention, on those social, economic and environmental factors which the intervention is designed to affect or may inadvertently affect.”¹⁶ Impact evaluations or assessments can produce several types of specific reports including poverty impact, regulatory impact, social impact, health impact, environmental impact and economic impact evaluation reports. “In the context of sustainable development, the social, economic, and environmental impacts of an intervention are all interlinked. The various types of impact assessment may therefore need to be combined in an integrated impact assessment, whose nature will vary according to the type of intervention, and the aims and cost-effectiveness of the overall impact assessment package.”¹⁷ Donors, governments and policymakers want to ensure programmes and projects maximise the potential output from scarce resources and consequently they are turning increasingly to the use of impact evaluations.

Impact evaluations or assessments, however, might not be feasible in all situations: there could be a reluctance amongst government officials and/or institutions to carry out impact evaluations because of political reasons (findings may be politically sensitive), cost considerations (deemed too expensive), a change in administration (there is a strategic shift in policy),

programme to their status before the intervention is a “before-after” comparison and ignores the impact of all the other factors that affect the beneficiaries. In order to separate the impact of the social transfer programme from all the other effects, it is useful to identify a group of households as similar to the beneficiary households as possible – and to monitor their living standards simultaneously with those of the beneficiaries. This group is called a “control group” because it enables the study to control for all the other factors affecting the households.

If the study is properly constructed, and if the beneficiary group is similar to the control group before the social transfer programme is implemented

– then differences between the beneficiary group (also called the “treatment” group) and the control group represent the impact of the social transfer programme. This is sometimes called “with-without” analysis – because the beneficiary group (“treatment” group) represents households “with” the intervention, and the “control” group represents households “without” the intervention. “With-without” assessments are generally considered more reliable than “before-after” evaluations.

SOURCE: Maluccio and Flores (2004).

technical complexity (resources and qualified personnel may not be available), required data may not be available, and the length of time it takes to do the evaluation may be too long. In addition the timing of the results may be sensitive (e.g. just before an election) or the results may be available too late for policy intervention.¹⁸

If any one of the following three questions can be answered positively, then doing an impact evaluation or assessment may be useful:¹⁹

- Is the programme of strategic relevance for national public policy?
- Can the evaluation results influence future design of the programme and other programmes?
- Will the evaluation contribute to improving the implementation or development of the programme?

A key reason why a formal impact evaluation may take place is to gain political support for the programme. For example, the Mexican government paid for the evaluation of Progresá in part because the conditional cash transfer model was relatively new and was viewed as a potential replacement for certain subsidy programmes. Support both from within government and among the public was needed to expand the approach.²⁰ Box 15.2 discusses the potential pitfalls of government mandated evaluations for social transfer programmes.

In a food security evaluation for Bangladesh, researchers identified three important features that enhanced the evaluation’s ability to improve public policy: (1) the research provided quantitative indicators and results, specifying concrete courses of action; (2) the evaluation team worked closely with policymakers; and (3) the results of the evaluation were provided in a timely manner that met the needs of decision-makers for information.²¹

Impact assessments should evaluate the reduction in poverty after the transfers have taken place, compared to the situation if the transfers had not been made. This requires identifying and evaluating a **counterfactual** situation – that is, how would participants have performed in the absence of the programme, and how would non-participants have fared if they were

Box 15.4: Quantitative and qualitative approaches for impact evaluation

Both quantitative and qualitative approaches are useful for effective impact evaluation. Several common approaches are described in this box.

Quantitative approaches

- Experimental design involves the random assignment of individuals or households either as beneficiaries, or as a control group which does not receive the service or good being provided by the intervention. This is also known as the experimental method. These evaluations can last from one to five years and range widely in cost from US\$50 000 to over a million dollars – depending on size and complexity. (This technique is sometimes referred to as randomised design.)
- Quasi-experimental design involves the use of a control group to match as closely as possible the characteristics of the beneficiaries receiving the intervention – either through propensity score matching or using a multivariate regression approach. This method often involves the use of large-scale sample surveys, and sophisticated statistical analysis. Like experimental design evaluations, quasi-experimental design evaluations can last from one to five years and range widely in cost from \$50 000 to over a million dollars – depending on size and complexity.
- Ex post comparison of project beneficiaries with a control group. With this method, multivariate analysis (to statistically control for differences in groups) may be used. This is one way of estimating the counterfactual situation. The cost

able to participate? The analysis calculates the average impact of the group who received the benefits from social transfers and compares this to that of a control group who did not receive transfers.²² It is also useful to compare the same individual or household over a period of time. “The critical objective of impact evaluation is therefore to establish a **control group** or a **credible comparison group** of individuals who in the absence of the programme would have had outcomes similar to those who were exposed to the programme.”²³ Box 15.3 provides an example of how the inclusion of a control group in a study can substantially alter the interpretation of the study’s results.

There are several methodologies for evaluating programme impacts using **quantitative data** and evaluation techniques. Two broad categories of design can be identified as **experimental design** (control group) and **quasi-experimental design** (credible comparison group).²⁴ **Qualitative** and **participatory** methods can also be used to assess impact. Such techniques enhance the quantitative methods and provide critical insights into understanding the viewpoint and perceptions of beneficiaries, the value of programmes to communities and towns, the underlying processes that may have affected outcomes, and generally provide a deeper understanding of the results observed in the quantitative analysis.

Understanding the context and variables can often help explain why two identical projects may have different outcomes in different communities. Several broad categories of contextual variables can be identified.²⁵ These include: economic factors, political factors, organisational or institutional factors, environmental factors and socio-economic or cultural factors. A technique often used by evaluators to capture this kind of information and provide ways to understand the linkages between these factors is called **Logical Framework Analysis** (sometimes called LogFrame Analysis). It is

is often less than half that of the experimental and quasi-experimental designs.

Combined quantitative and qualitative approaches

- Rapid assessment or review (conducted ex post). This method can encompass a range of tools to assess impact, such as participatory methods, interviews, focus groups, case studies, mini survey, direct observation and available secondary data. It generally uses quick low-cost ways to collect information. Since a main aim of rapid assessment is cost-effectiveness, evaluations with this design usually start from about \$25 000 and cost much less than conventional evaluations.

Qualitative approaches

- Participant Observation involves field researchers spending an extended amount of time in residence with a programme community, employing qualitative techniques and small-scale surveys. The technique often includes stakeholder analysis, participatory appraisal and beneficiary assessment.
- Case studies involve detailed or broad studies of a specific intervention involving open-ended questioning and the recording of personal stories.
- Participatory Learning and Action involves a facilitator assisting the active involvement of those who have a stake in the programme.

SOURCES: Adapted from World Bank (2004c) and Kirkpatrick and Hulme (2001).

a management tool mostly used at the project level, to improve the design of the intervention. “It involves identifying strategic elements (inputs, outputs, outcomes, impact) and their causal relationships, indicators and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of a development intervention.”²⁶

Experimental and Quasi-experimental design methods are often called **rigorous or quantitative impact evaluation methods**. These two methods have a strong advantage over the other methods as they are the most reliable: 1) for establishing causality, i.e. the relationship between a specific intervention and its actual impacts, and 2) for estimating the magnitude or size of impact that can be attributed to the intervention. They are able to distinguish the impacts of the intervention from the influence of other, external factors. These methods establish credible control or comparison groups. The superior strength of an experimental or randomised approach is that it helps ensure that the measured impact is only due to the intervention being evaluated.

However, randomisation is not always feasible or appropriate. Excluding someone from an intervention which could save his or her life because there needs to be a credible control group has very serious moral and political implications. Fortunately, researchers have been able to devise rigorous quasi-experimental design methods to compensate for some ethical considerations. Some researchers also make use of other, more **qualitative review methods**. Such methods will help address the views of key stakeholders concerning project operation and impact, and help to clarify and explain issues relating to the results chain or the causality relationships of the intervention. A mixed method approach to impact evaluation will also help achieve triangulation (provide two or more ways to understand and explain) evaluation findings.²⁷ Box 15.4 compares quantitative and qualitative approaches.

Box 15.5: Challenges with evaluations requiring experimental design

Many managers of social transfer programmes choose experimental design methodologies to provide robust and rigorous evaluations. These techniques, however, pose significant challenges for evaluators and policymakers.

Challenges:

- The random assignment of benefits to participants can be unethical.
- The randomness of the methodology sometimes creates political difficulties.
- Individuals in control groups are usually not passive subjects – denied benefits, they may seek social protection elsewhere – making it

difficult to interpret the results.

- Programme administrators have an interest in the evaluation's outcome – so they may intentionally bias the selection to improve the results.
- Experiments are often expensive and time-consuming.

Addressing the challenges:

- If there are budget or information gaps that make it impossible to effectively target the poorest eligible beneficiaries, the transparent and random assignment of participants to control and treatment groups can mitigate

Rigorous methods also have other disadvantages. One disadvantage is that they are usually expensive – rigorous impact evaluations often cost from US\$200 000 to US\$900 000, depending on project size, complexity and data requirements.²⁸ A second disadvantage is that they are often time-consuming, taking up to 2 years or more to complete; moreover, it can often take several years after a project has been in operation before its impacts start to emerge. These timing issues reduce the utility of rigorous impact evaluation when decision-makers require information quickly. Thirdly, rigorous methods are quite demanding in terms of the skills needed to conduct them; they require strong technical skills in social science research design, management, analysis and reporting. Finally, a criticism often levelled at impact evaluations is that they are unable to answer some important policy questions such as: how might the impact change if the programme design were changed?²⁹

Despite all of these disadvantages with the choice of method, the overall advantages of conducting impact evaluations still make it a particularly valuable (but costly) tool for long-term studies and research. The costs reinforce public good arguments for the international development community to help fund such evaluations, thus contributing to what is becoming a growing library of impact evaluation findings, available to all countries and to the donor community.³⁰

Randomised designs

Randomised designs are also known as experimental designs and are considered to be the most rigorous and robust of the evaluation methodologies. Box 15.5 identifies some of the challenges with evaluations that employ randomised design.

In a true randomised experimental design, individuals or households are randomly assigned (maybe through use of a lottery) to the experimental and control groups. If the sample is reasonably large the two groups can be assumed to be quite similar at the start of the experiment. The two groups

the negative consequences. A public forum for allocating participants with a demonstrably random process can offset some of the negative political consequences, although at the risk of highlighting the random character of the experiment.

- The incorporation of control group participants into the programme once the evaluation is complete can address some of the ethical issues in depriving them of social transfers during the evaluation stage. The random assignment of potential beneficiaries into the control and treatment groups then determines the timing of benefits – not whether or not

participants receive benefits.

- To address the ethical issues of denying benefits to the poorest, the design can randomise the benefit only to those participants above a minimum threshold of poverty – everyone below that level automatically receives benefits. The results of the impact assessment only strictly apply to the group for which participation was randomised, but this provides valuable information without compromising the needs of the poorest.

SOURCE: Baker (2000), pages 2–3.

are measured on indicators of the output or outcome that the experiment wants to produce. The intervention is then administered or provided to the experimental group and not the control group. The main benefit of this technique is the simplicity in interpreting the results. The programme impact on the outcome being evaluated can be measured by the difference between the means of the samples of the treatment group and the control group.

Often cited examples of experimental design evaluation include *Progresa* (a programme designed to increase school participation in Mexico), school-based health programmes in Kenya and India, and a school voucher programme in Colombia.³¹

When *Progresa* was launched in 1998, Mexican government officials made a conscious decision to take advantage of the fact that due to budgetary constraints they could not immediately roll out the plan to all 50 000 possible participant communities. They chose to begin the programme in 506 communities. Half of these communities were randomly selected to receive the programme and baseline and other data were collected in the remaining communities. The randomised phase-in of the programme allowed such clear documentation of the positive effects of the programme that even though the Mexican government changed hands, *Progresa* was maintained and even expanded to urban communities (*Oportunidades*).

Experimental designs are more difficult to apply in the real world. They require high skill levels and large resources for evaluation. However, there are many examples of advances in this form of design that make it more acceptable to use.³² Box 15.6 describes how experimental designs can be instrumental in overcoming the problem of bias.

Quasi-experimental designs

Experimental designs may be the optimal method for evaluating impact, but as mentioned before there may be several reasons why it may not be feasible to use such a rigorous method. “If random assignment is ruled out, it is still

Box 15.6: To experiment or not: the problem of bias

Impact assessments must evaluate all of the factors – social, economic, political, personal – that affect the success of social transfers. To control all of these factors, many methodologies compare the group receiving the benefits with a similar group that does not – the “control” group. It is important that the two groups be as similar as possible. If there are systematic differences between the beneficiaries (the “treatment” group and the control group), two types of errors can result:

- **Differences in observable characteristics.**

If the two groups are observably different from each other – in terms of average age, income, education level – then the two groups are likely to perform differently on performance

indicators that aim to evaluate the social transfer programme. It will be difficult to identify what causes the differences in performance – the programme, or the initial observable differences between the two groups. To address this problem, it is important to construct the two groups so that they are as observably similar as possible.

- **Differences in unobservable characteristics.**

Some differences between the two groups cannot be observed but nevertheless affect how individuals and households may participate in the programme. For example, more ambitious individuals may seek out the programme as a way to improve their livelihoods – and

possible under certain conditions to estimate impacts reliably using non-experimental methods.”³³ This group of techniques is called quasi-experimental (non-random) or non-experimental design. This method uses statistical tools and other econometric measures to generate a comparison group that closely resembles the group receiving the intervention.³⁴ Quasi-experimental design techniques can be further subdivided into two groups: a) techniques that address the bias caused by observable characteristics such as matched comparison methods and b) the techniques that address the bias caused by unobservable characteristics such as reflexive comparisons, double difference and instrumental variables.³⁵ See Box 15.7 for more detail on these techniques.

The most complete impact assessments evaluate both the effects of the social transfer programme as well as operational aspects. This involves both analysing the impacts of the programme as well interpreting the outcomes and their policy implications.³⁶

Impact assessments rarely answer all the questions that policymakers ask. Since they are based on evaluating actual processes and their impacts, it is difficult to answer the important “what if” questions. What would be the impact if the programme were scaled up to a national level? What if the programme were universal instead of targeted? In these cases, econometric techniques and micro-simulation models can provide important insight.³⁷

Careful design and planning enable impact assessments to answer many different types of questions.³⁸ For example, an evaluation of a conditional cash transfer programme with conditions for school attendance and healthy clinic visits might address the following questions:

- Do students participating in the programme attend school more frequently?
- Does the programme reduce the poverty of participating families?
- Does the nutrition of participants improve?
- Which age groups or household types benefit the most?

also succeed better in the face of other opportunities. As a result, the treatment group may perform better on a wide range of indicators compared to a control group that is made up of less motivated individuals. An assessment that ignores this effect is more likely to attribute positive performance to the social transfer programme – and not to the unobservable characteristics that may be responsible. This kind of error is called “selection bias”.

One way to eliminate these types of error is to randomly assign eligible participants to either the control group or the treatment group. Since

individuals cannot select which group they join, there is no apparent selection bias. If the groups are large enough, the random assignment will tend to make them similar on average.

These experiments can be expensive and sometimes controversial. They require substantial planning, from the time the social transfer programme is first conceptualised. Non-experimental designs may be less expensive – and simpler practically and politically. However, they involve greater conceptual challenges in terms of interpreting the results.

SOURCE: Blomquist (2003), page 12, adapted from Baker (2000), Ravallion (1999) and Orr (1999).

- What characteristics contribute most to the success of the programme?
- Is the programme cost-effective?
- Are the programme benefits greater than the costs?³⁹

Lessons of impact assessment

Economists and social policy analysts have identified a number of important ingredients for good impact assessments. The following highlights a number of key issues that practitioners consider.⁴⁰

- The foundation for a good assessment is a thorough understanding of the programme, in particular the administrative and institutional details.
- It is critical to understand the social context, including the geographic distribution of poverty (for example, as depicted by a poverty map), the workings of the labour market, ethnic and other demographic characteristics of poverty, and the policy context – including other relevant social programmes.
- Be open-minded about your source of data. Scientifically sampled survey data is excellent – but additional and as important information might be gleaned from interviewing programme participants, managers and staff.
- Be careful about simply comparing outcomes for programme participants and non-participants. For example, in a particular rural village, a randomly selected group of non-participants may have lower poverty rates than a similarly randomly selected group of programme beneficiaries. That cannot support the conclusion that the programme is ineffective – the programme should be drawing on households with higher poverty rates in the first place. Good comparisons must control these important differences between participants and non-participants.
- Both quantitative and qualitative data can be useful, as long as it is organised appropriately. It is often helpful to link survey data to more specific information about the programme and geographical setting.⁴¹

Box 15.7: Quantitative methods for assessing the impact of social transfers

Several of the main quantitative methods for impact evaluation are discussed in this box. Because no single method is always appropriate, it is often desirable and appropriate to combine techniques.

Experimental designs (randomised control)

- **Experimental design** is one of the most rigorous methods for evaluating impact. The technique randomly selects participants into “treatment” and “control” groups. The “treatment” group receives the social transfer and the “control” group receives nothing. The aim is for the two groups to be as similar as possible so that the only difference is the receipt of the social transfer (or other “treatment” that is being tested). By randomly assigning the eligible participants to the two groups, there is no choice on the part of the experimenters or the potential beneficiaries that can bias or distort the results of the evaluation. If the test is structured properly, there will be no expected difference between the control and treatment groups except for the impact of the social transfers programme on the treatment group. However, because of random differences between the two groups – referred to as sampling error – there may be differences in impact that cannot be attributed to the programme. The larger the number of participants in the two groups, the smaller is

this kind of error. Experimental designs are sometimes controversial because they may involve denying people access to social security.

Quasi-experimental designs (non-experimental)

- **Propensity score matching.** Because experimental design is not always practically, politically or ethically appropriate, alternative approaches aim to achieve similar results by creating control groups from the population that did not have access to the social transfer programme. The techniques use “matching methods” to construct control groups by selecting the non-recipients in a larger survey that best match the characteristics of the beneficiaries in the treatment group. One commonly used technique is called “**propensity score matching**”, which matches the treatment group beneficiaries to non-recipients from the larger survey using a score (the “propensity score”) calculated from observable characteristics assumed to be relevant for defining individuals for the purpose of the impact assessment. The propensity score aims to predict the probability an individual or household would participate in the social transfer programme. A good control group has a distribution of propensity scores similar to that of the treatment group – but the treatment group receives the social

Implementing an evaluation of a social transfer programme

The previous sections of this chapter provide an overview of several important methodologies useful for evaluating social transfer programmes, as well as a discussion of important design issues for impact assessments. In practice, there are a number of steps required to effectively implement an evaluation in order to improve the delivery of social transfers.⁴² (Box 15.8 discusses several of the cross-cutting issues that can improve a results-based approach.)

STEP 1: Deciding to evaluate

The first step is to determine the appropriate timing for the evaluation. As discussed previously, impact evaluations involve considerable financial costs and political risks, yet potentially may yield important strategic, practical

transfers while the control group does not. In constructing a good control group, it is useful to draw members from the same social and economic environment as those in the treatment group – and to survey them using the same questionnaire and with similarly trained interviewers.

- **Double difference.** One of the most important analytical techniques for analysing the resulting data is called the “double difference” (or “**difference-in-differences**”) methodology, which compares the treatment and control groups both before the social transfer programme begins, and again during the implementation of the programme (possibly multiple times). The “first” difference is how the treatment group varies from the control group, and the “second” difference is how these differences vary before and after the programme has been implemented. Sometimes households in the control group with characteristics vastly different from the average (called “outliers”) are dropped from the comparison – particularly if their propensity scores are outside the range observed for the treatment group.
- **Instrumental variables** is another important methodology which uses statistical techniques to control for participation. The key is the identification of one or a set of variables

that will affect an individual or household's participation, but will not affect the impact of the programme given that the individual or household is participating. This is often very difficult and the source of much of the judgement involved in this kind of methodology. When the analysis is carefully constructed, the technique identifies the impact on the participants that is attributed to the programme. In the process, the technique calculates a formula that predicts the probability that an individual or household will participate in the programme – based on the values of the “instrumental variables” – in a manner similar to that for propensity test scoring.

- **Reflexive comparisons** involve a baseline survey of beneficiaries prior to the implementation of the social transfer programme, followed by another survey once the transfers have had an opportunity to make an impact. The technique compares indicators from the baseline to the follow-up survey. This technique may mistakenly attribute impact to the programme when in fact it resulted from other changes in the economy or society occurring at the same time as the social transfer programme.

SOURCE: Baker (2000), page 7–8.

and political benefits. If the programme is still relatively young, it may not be sufficiently developed to demonstrate measurable results. More innovative and replicable programmes yield potentially higher benefits from evaluation and are more likely to justify the cost and risk.⁴³

STEP 2: Making the objectives clear

The decision to proceed will weigh the objectives of the evaluation and the likelihood of achieving their success. The second step is to clarify these objectives, balancing a comprehensive scope with the specificity required to achieve success. As part of this step, it may be useful to establish a reference group including representatives of all the key stakeholders. This will help ensure that all the critical questions are reflected in the statement of objectives, as well as facilitate the reflection of the findings in improved programme delivery once the evaluation is complete.

Box 15.8: Linking impact evaluation with a results-based approach

Increasingly governments and international development agencies are emphasising the need for programme evaluation. This trend reflects greater demands for improved accountability by service providers – particularly to their funders – as well as an increased focus on results-based management that aims to improve programme efficiency.

Unfortunately, impact assessments are often under-funded, in spite of evidence that these investments can generate substantial social returns. Many of the benefits of sound evaluations accrue to global community of social policy analysts – and the social transfer programmes they support – and efficient mechanisms do not exist to compensate evaluators for the public good they produce. In addition, programme managers are often reluctant to allocate scarce resources for evaluation purposes – the results are often available too late to make the maximum impact on the programme, and the focus of evaluations frequently fails to identify recommendations for operational improvements.

Some particularly frank politicians acknowledge that they have little enthusiasm for funding evaluations because their electoral time horizon

is too short to benefit from the longer time lags necessary to generate productive assessments. The long horizon of policymakers responsible for Mexico's *Oportunidades* (previously *Progres*a) has improved the payoff from evaluation and increased the public's awareness of the successes of the programme – ensuring the sustainability (if not the name) of the social transfers system even through changes in presidential administrations.

Three reforms can improve the capacity of programmes to benefit directly from evaluations – and thus improve prospects for appropriate funding of impact assessments.

1. **Just-in-time delivery of results.** Evaluations should deliver short-term results to managers while the programme is active. It is important to guard against developing a short-term mentality that undermines longer term gains, but there are substantial benefits from providing early feedback that supports programme improvements.
2. **Results-based management.** Impact evaluations that ensure accountability require impartial external auditors. Assessments that

STEP 3: Identifying the evaluator

The third step is the identification of the evaluator, and as the subsequent steps unfold, the formation of the evaluation team. Credibility requires that the evaluator demonstrate independence from the institution or agency responsible for the social transfer programme. Often an implementing agency will identify the evaluator based on competitive bids. Box 15.9 identifies several of the key components of the terms of reference required for a request for proposals. Depending on the capacity of the programme team that is requesting the bid, part of the evaluation design may be included in the request for proposals. Alternatively, the request for proposals may require that potential bidders propose an appropriate methodology.

STEP 4: Fully designing the evaluation

The fourth step involves the full design of the evaluation, including the definition of a rigorous methodology. The identified evaluator may put this together as part of the proposal process – or in some cases the programme's institution may set forward the required methodology as a precondition for the bid. Section 15.2 of this chapter discusses some of the most important design questions that this step should address. Once the evaluator is confirmed, the programme's institution and the evaluation team must thoroughly review the

aim to improve operational results, however, benefit from participation by managers and staff. Evaluations are learning processes, and they must harness the information from all members of the organisation. It is challenging to integrate the accountability and results-based objectives of evaluation – but meeting this challenge provides many more benefits than does a conventional evaluation.

3. **Experimenting.** One can further improve results by experimenting in order to identify better ways to achieve programme objectives. It is important to identify the drivers of performance – such as targeting rules, the determination of the size of transfers, delivery mechanisms, complementary programmes and mechanisms for accountability. Experimenting with alternative approaches can inform management of better ways to achieve improved results.
4. **The value of administrative data.** Most social transfers programmes maintain rich databases capturing information about beneficiaries – this is referred to as “administrative”

data. Administrative data can sometimes provide information about participants over time – if the databases are updated with current beneficiary information. Access to administrative data is usually impossible for researchers unaffiliated with the implementing institution – but if the evaluation is commissioned by the programme organisation, the evaluators should consider the potential contribution of administrative data.

5. **Community politics.** When local communities are actively involved in the implementation of programmes, particularly with respect to targeting, a great deal can be learned by comparing how different localities address the same issues and problems. Understanding the local politics – community decision-making, political trade-offs, the exercise of influence – is essential in order to effectively evaluate the social transfer programme.

SOURCE: Sadoulet (2004).

proposed methodology. This should include a basis for comparing the impact of the social transfer programme to the counter-factual (employing some type of control group methodology). Matched comparison groups and experimental design provide some options for consideration, in light of the issues discussed above. The confirmation of the methodology also requires decisions on budget. Box 15.10 discusses how methodological issues – particularly with data requirements – can affect budget requirements. Since some bid protocols require a full costing with the proposal, these budgetary issues might need to be addressed in step 3.

STEP 5: Mobilising the data

Data comprise the lifeblood of the evaluation. The starting point of mobilising the necessary data is determination of the appropriate sample size (with its cost implications). The sample needs to be sufficiently large to detect significantly the impacts of the programme. The finer the granularity of the sub-groups for which measurable effects are desired, the greater will be the required sample size. For example, drawing inferences about the national impact will require less data than identifying and comparing impacts at a provincial or state level. Often the data phase of the evaluation is one of the most expensive components (perhaps second to analysis – in some cases greater).⁴⁴

Box 15.9: What to include in the terms of reference for a request for proposals

When commissioning an external impact evaluation or assessment a short brief or terms of reference should be drawn up to provide the evaluators/consultancy with all the relevant information concerning your needs for the evaluation. This will allow them to put together an informed tender or bid. Such an evaluation brief is a short document – about four to six pages in length – that outlines the scope of the evaluation. The document should include:

- **Project title and brief description**

- **Background**

This should provide general background information and reasons for the evaluation study.

Who is commissioning the evaluation – the organisation being evaluated, another agency or a funder?

Provide brief summary details about the programme being evaluated. For example, what service is being provided, how many staff members are involved, who are the recipients of the service, appropriate materials such as marketing leaflets and a website address.

Provide information about the purpose, focus, use and audience of the evaluation. For example, will the whole organisation or a programme be evaluated? Is the purpose of the evaluation to help strategic planning, get more funding, or evaluate impact? How will the evaluation be used? Is it to demonstrate accountability to funders or improve the service? Who will read and use the evaluation report?

Outline issues that may affect your organisation's view of the evaluation process. You may want to

use a participatory approach in your evaluation methods.

Outline issues related to data collection. What data already exists, what difficulties arise in collecting new data?

Provide information about internal responsibilities and liaison. Who is the main contact person on this evaluation? What work will be carried out internally?

- **Aim and objectives**

Document the main single aim and the objectives, and how and who will use the impact assessment.

- **Specify the assessment's scope and methodology**

- Specify the work requirements for the evaluation team.

- Identify the main research questions.

- Indicate relevant partners and stakeholders for the impact assessment.

- Identify the appropriate methodology for the assessment.

- **Outputs**

The terms of reference should specify the general outcome and expected deliverables. Each activity should be associated with an expected output. Include consultation of the draft report with stakeholders to reflect their role in the process. Also include dissemination of the report as an outcome.

- **Required experience, competency, qualifications, and expertise**

The terms of reference should indicate the minimum requirements and request appropriate documentation, including résumés, CVs,

It may be possible to use existing national surveys if they capture sufficient information about the programme. In many cases new data must be collected – requiring the development of a survey instrument, hiring and training fieldwork teams, pilot testing the survey tools, the fieldwork itself (including validation), the design of the appropriate databases and tools, and the entry and cleaning of the data.

Increasingly social transfer programmes begin the evaluation process prior to the implementation of the programme – so this phase begins with the collection of baseline data. Subsequent follow-up steps collect additional waves of data in order to track the impact of the social transfer programme over time.

capability statements, description of country knowledge, and other appropriate information.

Specific requirements include:

- How does their professional expertise relate to the subject of the assessment?
- What is their evaluation experience, including experience with methodologies?
- What is their knowledge of the country or region?
- What is their previous experience with this specific type of assessment?
- What is their experience and awareness with gender and cultural issues?
- Do they possess the necessary language skills?

- **Conduct of the work**

This should set out clearly how the project will be implemented:

- The role of the evaluators – both the team leader and team members
- The design, implementation and target dates of the work programmes,
- Specify how the local counterpart personnel will participate in the process.

- **Specify your reporting requirements**

Clearly indicate the schedule of progress reports as well as the deadlines for the final report. In particular, carefully document:

- A schedule for the timing of the work, including an indication of any flexibility that is available.
- A full schedule of progress reports, interim and final reports, and requirements for result reporting.
- What kind of final report is required, and will this include a formatted presentation?
- Specify intellectual property ownership –

who will own the rights to the work and the copyright?

- **Dissemination of findings**

Depending on the requirements of the project, the needs for dissemination can be planned and included in the terms of reference so that the budget contains sufficient resources. Reports or summaries should be translated into local languages as appropriate and costs for this included in the budget. Several methods of dissemination can be considered:

- The results can be presented at conferences or in workshops or seminars
- Formal reports presenting the results of the impact evaluation, as well as briefs and summaries
- Academic articles and more accessible stories in appropriate publications.

- **Budget allocation**

If appropriate, indicate your available budget allocation for the assessment, and whether it includes VAT.

- **Timing of the call for proposals**

What is the submission date required for the proposals? It is useful to specify if you plan to have any arranged interview dates. Give details about the person to contact for any clarifications or further information.

SOURCE: draws from Charities Evaluation Services (2004).

STEP 6: Analysing the data

A plan for data analysis should precede the commencement of fieldwork. Once the survey is in the field, the range of possible analyses will face constraints imposed by the data. It is critical that the required analysis be conceptualised hand-in-hand with the design of the data requirements.

The evaluation should achieve more than the identification of programme effects. Analysis of programme processes can identify operational shortcomings and proposals for improved cost- effectiveness. In addition, qualitative analysis and more sophisticated data analysis can shed light on the causal reasons for the observed impacts.

Box 15.10: Budgeting for impact evaluation

The following elements determine the cost of an impact evaluation. It is useful to include detailed specifications for these steps in the evaluation's request for proposals so that the cost of the project can be accurately estimated.

- **Collection of baseline data**

The most effective evaluations require baseline data collected prior to the implementation of the social transfer programme. The costs depend on the type of data required, how it is collected, and by whom. The availability of skilled local enumerators – and their training requirements – also affects the cost. Likewise, the need for external consultants can significantly affect budget requirements.

- **Impact evaluation design**

The design of the impact evaluation requires thorough consultation with the stakeholders in order to ensure a comprehensive approach. The cost of this component will depend in part on these meetings – how many, who is involved, the costs of participation and the need (if any) for external facilities. The type of impact analysis selected also determines the budget required for the evaluation – balancing the costs of various options against the benefits in terms of better information, considering the opportunities to train local staff to carry out the work.

- **Regular data collection**

Impact evaluations require an ongoing process of collecting data that measures the programme's impact. In some cases, existing governmental statistical surveys may provide sufficiently

STEP 7: Reporting the results

The final step for the evaluation team is the writing up of the analysis, conclusions and policy recommendations and the reporting of these findings to the programme managers, policymakers and other stakeholders – including the reference group if established in step 2 above. In many cases, the reports are not available for one to two years after the final wave of data collection. The evaluation plan should carefully (and generously) schedule milestones and deliverables to reflect the decision-making calendar of policymakers.⁴⁵ Distinct forms of the report will target different audiences. A rigorous technical analysis is the bedrock of the evaluation. Clear and accessible presentations maximise the impact on policymakers. Concise and attractive policy briefs can broaden the policy impact and provide a vehicle for communicating the value of the programme to a more expansive audience. Making the reports and datasets available on the Internet can nurture important external benefits – creating public goods for the rest of the world and perhaps eliciting further independent studies that provide additional information for the benefit of the programme.

STEP 8: Reflecting the results in improved programme delivery

The decision in step 1 to go ahead with an evaluation should reflect a plan for using the results to improve the delivery of programme objectives. Evaluations that simply meet legislative or donor requirements are unlikely to justify their costs. Keeping the reference group (established in step 2) informed of progress through periodic workshops can lay the foundation for more effective results by contributing to the stakeholders' participatory ownership of the evaluation. Mexico's *Oportunidades* and South Africa's comprehensive system of social grants provide two examples (certainly among many others) of programmes

detailed data. There may be significant analysis required to process the existing data into a format usable for the evaluation. In other cases, the evaluation will require quantitative or qualitative surveys to generate the information necessary for the assessment. The cost will depend on the magnitude of this task and the availability of skilled local staff or other personnel who can carry out this work.

- **Analysis of information**

The heart of the impact evaluation is the analysis of the information. The cost will depend on the technical requirements from the design phase, which determines the level of skill and experience required. The duration of the programme and the number of progress reports required will also have a significant impact on the

necessary budget. The availability of local staff or external consultants will also affect the cost.

- **Reporting and disseminating the findings**

The impact evaluation produces a final report as its key output, which must be effectively disseminated. The cost of this component depends on the reporting requirements and the extent of dissemination. Translating the report into local languages may be necessary for it to have the desired impact. Workshops may be useful to convey the results and involve the stakeholders. Who should receive the report and who should participate in the workshops? Is it useful to publish articles in journals and present the results at conferences?

SOURCE: Kirkpatrick and Hulme (2001), page 17.

that have continuously used evaluations to improve and expand the delivery of social transfers.

Conclusions

This chapter highlights the importance of an eclectic approach to monitoring and evaluation. A variety of methodologies serve the diversity of objectives policy analysts seek to meet. Good assessments are rooted in a thorough understanding of the social transfer intervention and the context of poverty it addresses, and the groundwork for monitoring and evaluation begins with the initial design documents for the programme. Quantitative and qualitative techniques complement each other, and rigorous analysis must grapple with the omnipresence of the potential for error.

The illuminating power of effective monitoring, evaluation and impact assessment provides the essential evidence linking programme performance to improvements in design and implementation. Positive evaluations can mobilise political support and expand the resources available for scaling up scope and coverage. Monitoring and assessment can identify problems and propose solutions, and the evidence generated can serve not only the programme itself but also the global community involved in designing and implementing social transfer programmes.

Endnotes

- 1 Blomquist (2003), page 3. See also Ezemanari et al. (1999), page 1.
- 2 Kusek and Rist (2004), page 227.
- 3 Blomquist (2003), page 3.
- 4 Adapted from Kusek and Rist (2004), page 225.
- 5 Kusek and Rist (2004), page 117.
- 6 Blomquist (2003), page 3.
- 7 Baker (2000), page 1.
- 8 For more detail on different evaluation approaches and strategies, see Chen (2005), page 48; Bamberger et al. (2006), page 46.
- 9 Baker (2000), page 7. See also Duflo and Kremer (2003).
- 10 Bamberger et al. (2006), page 19.
- 11 Baker (2000), page 1; Blomquist (2003), page 3.
- 12 See Bamberger et al. (2006), page 27 and chapters 3 and 4 for more detailed information on these.
- 13 Blomquist (2003), page 4.
- 14 Baker (2000), page 1.
- 15 Charities Evaluation Services (2004), page 2.
- 16 Kirkpatrick and Hulme (2001), page 2.
- 17 Ibid.
- 18 Baker (2000), page vi; Blomquist (2003), pages 1–2.
- 19 Blomquist (2003), page 19.
- 20 Ibid., page 22
- 21 Ibid., citing Babu (2000).
- 22 Baker (2000), page 1.
- 23 Duflo and Kremer (2003).
- 24 See Box 7.4 for more detail.
- 25 Bamberger et al. (2006) identifies these categories and provides more detail (page 42).
- 26 Kusek and Rist (2004), page 226.
- 27 Bamberger (2006), page 27 and World Bank (2004f).
- 28 World Bank (2004f), page 23; Blomquist (2003), pages 17–18.
- 29 Blomquist (2003), page 22.
- 30 Duflo and Kremer (2004).
- 31 Duflo and Kremer (2003).
- 32 Ibid., pages 5–18.
- 33 Blomquist (2003), page 8.
- 34 Baker (2000), page 8.
- 35 Blomquist (2003), page 8.
- 36 Ibid., page 3.
- 37 Ibid., page 4.
- 38 Ibid.
- 39 Ibid.
- 40 For example, see Ravallion (1999), Ravallion (2001). The following considerations draw substantially on the lessons illustrated and concluded in these papers.
- 41 Ravallion (2001), page 136.

- 42 This section consolidates and adapts the frameworks of Baker (2000, chapter 2) and Blomquist (2003, page 22), which in turn draws on Ezemanari et al. (1999).
- 43 Baker (2000), pages 17–18.
- 44 Blomquist (2003), pages 14–16.
- 45 Baker (2000), page 39.

