



Measuring cause, monitoring effect: International aid agencies can record how many sacks of food are delivered and whether the nutritional levels of children go up or down, but what is the true impact of humanitarian assistance? What is it doing to help reduce vulnerabilities and build more sustainable livelihoods? In a search for effectiveness, efficiency, professionalism and accountability, aid agencies today are increasingly trying to evaluate the impact of their assistance as well as monitor the process by which it is provided.

Food delivery, Mozambique, 1985 Jean Gaumy/Magnum

Evaluation: Measuring effects not process

There are three powerful reasons why the effectiveness and impact of relief programmes should be regularly assessed. First, they are intended to save lives and reduce suffering so their effectiveness is crucial to the affected population. Second, the international community spends hundreds of millions of dollars each year on relief assistance while governments and communities in affected countries spend hundreds of millions of dollars more. Third, evaluations can, potentially at least, fulfil an important accountability function. Financial accountability mechanisms for agencies and donors are well developed but mechanisms for ensuring accountability to the population being served are not.

Even a decade or two ago many within agencies regarded evaluation of relief programmes as unnecessary, if not unwelcome. The view was: "Our motives were well-intentioned, we did our best under difficult circumstances, why should we now subject ourselves to a critical examination – anyway, all disasters are different, so what is the point in trying to learn the lessons of our response to this particular disaster?" Fortunately, such attitudes are now less prevalent. Most bilateral and multilateral donor organisations set up evaluation units during the 1970s and, following the African emergency in the 1980s, began to include studies of their organisation's response to emergencies in their work programmes. In the 1980s, some larger international NGOs developed evaluation units to review their response to crises.

However, despite a shift in attitudes, the situation remains unsatisfactory. Many relief agencies still do not evaluate programmes. Even where agencies do systematically examine their own performance, it is rare for the aggregate response by all agencies (UN, donors, governments, NGOs) to be evaluated. Given the importance of relief programmes to affected populations, the scale of resources expended and the potential for accountability, it is hard to understand why so few

evaluations are carried out. This chapter argues that powerful factors discourage evaluation of relief programmes and limit evaluation efficiency when they do take place. Many factors stem from the nature of emergency programmes and the organisational, technical, methodological and political difficulties evaluators face.

Definitions and techniques

Relief evaluators look to development assistance evaluation methods for models of how to work, but the principal approaches used in evaluating development programmes are often inappropriate for relief. Most donor organisations have similar definitions of evaluation, such as that of the Organization for Economic Cooperation and Development (OECD): "An examination as systematic and objective as possible of an ongoing or completed project or programme, its design, implementation and results, with the aim of determining its efficiency, effectiveness, impact, sustainability and the relevance of its objectives."

Monitoring and evaluation are inextricably linked because evaluations are heavily reliant upon information gathered by monitoring systems. Casely and Kumar define monitoring as "a continuous assessment both of the functioning of the project activities in the context of implementation schedules and of the use of project inputs by targeted populations in the context of design expectations. It is an internal project activity, an essential part of good management practice, and therefore an integral part of day-to-day management."

Since the 1960s, when donor organisations first began to evaluate development assistance, a range of common evaluation techniques have been developed. In analysing efficiency, Cost-Benefit Analysis (CBA) has been a dominant approach. CBA requires an estimation and comparison of the costs and benefits to a society over a project's lifetime. Despite improvements, CBA has a number of limitations, particularly for projects with a diffuse re-

relationship between inputs and outputs or where there are important non-quantifiable outputs

Project impact analysis requires identification of any changes that have occurred since the intervention started and the nature of any causal links. A detailed description of the situation prior to the intervention is needed, and thus many development projects include a baseline survey phase in the project cycle. Ideally baseline studies will be designed to help later evaluation. In the absence of such studies it may be possible to "construct" a baseline by using other sources, such as reports and statistics, or through recall surveys of the beneficiary population, though such sources will often be less accurate and not tailored to the evaluators' needs.

Establishing the nature of causal links is difficult. Measuring before and after situations and attributing all changes to the project is unacceptable. A variety of techniques have been developed to separate the influence of "confounding factors" from the project's effect, including comparing the target population with control groups, interrupted time-series design in which the target population is systematically monitored before, during and after intervention, and econometric techniques to establish the statistical likelihood of a link between observed changes and potential explanatory factors.

Attempts have been made to develop inexpensive evaluation techniques which can better judge social development objectives and pay greater attention to the views of the target population. Many of these

attempts have been instigated by NGOs or designed to meet the evaluation needs of small-scale NGO projects with substantial social development components. Logical Framework Analysis (LFA) and Rapid Rural Appraisal (RRA) are now widely used not only by NGOs but increasingly by bilateral and multilateral donor organisations. LFA involves constructing a matrix with the project objectives, outputs and inputs listed on the vertical axis and indicators of achievement, means of measurement or assessment, and assumptions and risks on the horizontal. This does not constitute evaluation, but when used in project design it can greatly facilitate evaluation by encouraging clear statements of objectives, exposing the logic of how the project is expected to work and requiring identification of those indicators by which progress can be monitored and evaluated.

Rapid Rural Appraisal developed out of an awareness of the limitations of large-scale questionnaire surveys and the distorted views that could be gained by brief rural visits, so-called "development tourism", by urban-based professionals. RRA includes flexible, opportunistic, improvised and iterative approaches to learning rather than standard blueprints; techniques for offsetting spatial, seasonal, professional and personal biases often involved in less structured and sensitive approaches to information gathering. It also involves rural people in the learning process, cross-checks information from more than one source; optimises information collection to obtain only that which is directly useful; and critical self-awareness by

Box 5.1 Listening to the beneficiaries

Woldu Desta has not got a friend left in town. As a member of Adigudum's Relief Committee, he frequently has to tell farmers why they can't get any relief food. "You are just not poor enough", is the most common explanation in this Tigrayan community where only the absolute destitute receive aid.

"If people have one ox, donkey or camel they are not totally destitute. I often quarrel with farmers who say they will then sell their animal in order to qualify for the Ethiopian Red Cross Society (ERCS) distributions. I tell them not to, but mostly they will not listen to me," says Woldu.

Out of Adigudum's 52,000 inhabitants, only 4 per cent can be described as self-sufficient, 16 per cent are poor, possessing a small plot of land and maybe one ox. Fifty-six per cent are very poor, without any draught animals and often forced to let better-off farmers cultivate their

land in return for a third of the harvest. The remaining 24 per cent are destitute, having no land, no animals, and often unable to work as labourers. The ERCS, which distributes in this area, is only feeding the destitute and some of the poor, leaving many to survive on an absolute minimum. A tough but necessary policy in a country where millions are at the risk of starvation.

But Woldu Desta, 55, is among those who never received any relief food. He used to be one of the rich people in Adigudum until the big drought struck in 1984/85. Out of his 15 oxen and 110 sheep and goats only one small cow survived. During the past ten years he has never possessed more than four draught animals, and this year the number has once again been reduced to one; an old and tired ox that helped him harvest 400 kg of wheat from half a hectare of land during the past season. A fourth of the harvest has

since been used to repay loans, leaving Woldu Desta and his wife with barely enough to survive.

The eyes of Woldu Desta light up as he talks about the past. About drinking milk, eating meat, having butter and honey for breakfast. Now

adays, a chicken may be killed for special celebrations, but the daily menu consists of bread with a stew of water, spices and a few vegetables.

And while the farmers in Adigudum are struggling to get the most out of their scarce soil, many of their sons have left. They now work in other regions of Ethiopia, in Eritrea or Sudan, sending home money to sustain the family. A small migration is constantly taking place. Avoiding the larger one is among the challenges facing the community and the aid agencies.

Adigudum still remembers how 1,500 people were buried during one day of 1985. Things are better now. But not much. ■

those undertaking assessment. A further development is Participatory Rural Appraisal, which shifts primary responsibility for collation, analysis and presentation of information from "outsiders" to rural people.

These methods have three problems. First, evaluating development assistance usually required that project outputs could be readily identified, measured and valued. Second, approaches to the measurement of efficiency and impact are "data heavy" and costly in time. Third, the body of knowledge of development assistance evaluation is currently in a state of flux: "new" techniques have been developed to take account of shifts in the content and objective of development projects, NGOs have a growing role in development, and there is increased respect for the views of the population being served. Such techniques are only just beginning to be used in the evaluation of relief programmes.

Relief evaluation difficulties

There is often confusion when the term evaluation is used in relation to relief programmes. Strictly speaking, studies which do not effectively address all five aspects of evaluation – relevance, efficiency, effectiveness, impact, sustainability – should not be called evaluations. Because of the exceptional methodological and technical difficulties in assessing relief programmes, many "evaluations" do not or cannot assess the impact or the assistance provided and restrict themselves to assessments of relevance, effectiveness and efficiency. Even such restricted "evaluations" face numerous difficulties stemming from a lack of measurable objectives and inadequate data. Thus "review" would often be a more appropriate term for many evaluations of emergency programmes.

While private sector audits are generally associated with checking the probity of an organisation's financial management, public sector audits tend to have the broader objective of assessing conformity of an agency's actions to its objectives and authorised procedures. There is often substantial overlap between audits and evaluations. Audit assessments of particular emergency programmes or types of relief assistance have been undertaken by government audit offices such as the US General Accounting Office and the UK National Audit Office and, with the rapid increase in foreign aid resources for emergency programmes, such studies are likely to become more frequent. However, such audits are more restricted in their remit than evaluations, focusing on questions of value for money, appropriate tender procedures and the speed of assistance. Questions of the appropriateness of the response and its impact are invariably beyond the remit of such audits.

All relief programme evaluators face the lack of appropriate data. Accurate in-

formation on crucial indicators, such as excess mortality, is often unavailable. The reasons for the lack of data include severe time pressure, lack of thought about information that will be needed for evaluation, and ethical factors.

Time pressures

In the immediate aftermath of sudden-onset disasters there is the need to initiate search and rescue operations within the first few critical hours and to begin providing medical care, food, shelter and counselling to survivors. Even in slow-onset disasters, such as drought, the need to act quickly is only slightly less pressing as a result of the long lead times invariably involved in mobilising and delivering food aid. Food-needs assessment have to take place either during or immediately after the (reduced) harvest to provide sufficient time for distribution targets to be met before the next rainy season hampers the logistics operation. Even within chronic "complex" emergencies, events often require a rapid response, such as population displacements or sudden access to a conflict-affected area.

The need to act quickly has important implications for the sequence of activities and the time available for each one. A typical development project cycle involves these distinct stages:

- identification;
- design;
- appraisal,
- approval;
- the collection of baseline data;
- implementation;
- completion;
- operation (e.g., integration of project activities into those of local institutions);
- evaluation

From project identification to the start of project implementation takes, for many development projects, from 12 to 36 months. In a relief response to a sudden-onset disaster or sudden population displacement, the sequence is effectively "collapsed" by time pressures to:

- hazard event/population displacement;
- needs assessment;
- design of intervention/preparation of funding proposals;
- approval;
- implementation (relief phase);
- design of rehabilitation activities;
- implementation (rehabilitation phase),
- completion;
- evaluation

For international relief agencies responding to a sudden-onset disaster, implementation frequently starts within two to three days of the event, needs assessment forms the basis for project design, and the stages of appraisal and collection of baseline information found in a development project cycle are effectively

omitted. Because of the lack of baseline information, evaluators may have to rely on information gathered during the needs assessment to try to determine the relief programme's impact. However, it is rare for a needs assessment to contain information which is sufficiently detailed or comprehensive for it to meet the needs of evaluators interested in measuring the programme's impact. The principal purpose of a needs assessment is to determine the type and scale of assistance required, not to collect information as a baseline against which intervention impact might be measured.

Relief agency personnel themselves operate under severe time pressure. Staffing levels are rarely able to keep pace with the expansion of workload. Few relief agencies operate with sufficient surplus capacity to redeploy appropriately-skilled personnel at short notice and suitable personnel are rarely available at short notice. Over the years, registers have been set up of skilled personnel available at short notice, which now operate with varying degrees of success. Even where the supply of skilled personnel is not a constraint, other factors, such as restrictions on the number of expatriates involved in a programme or delays in obtaining entry visas and work permits, may prevent relief agencies from reinforcing existing personnel. Available staff invariably concentrate on those activities which directly address immediate problems. Information collection for a possible later evaluation is a very low priority.

Basic information which would take little time to collect or record is often unavailable to evaluators simply because there was a lack of thought about information that might be needed for monitoring and evaluation. Seed distributions fail to note dates so their timeliness for planting cannot be determined, harvest yields are not recorded which could determine the effectiveness of rehabilitation efforts, and agencies in receipt of donated trucks frequently do not record vehicle utilisation rates, information which is crucial to evaluators attempting to assess whether money might have been better used in other ways.

Large amounts of data collected during some relief operations goes unused simply because the need was not properly assessed before the decision was taken to collect it. One review of targeting methods employed by NGOs during the mid-1980s African emergency found that large amounts of data on the condition and coping mechanisms of affected populations were collected using scarce personnel resources but much of the information was never used by the NGOs concerned or made available to researchers.

It is not uncommon for agencies to undertake two or three surveys of an affected population over a period of time, but for each survey to use different sampling techniques, because of staff changes

or the particular purpose of each survey. These cannot be directly compared, depriving evaluators of the chance to construct a longitudinal "picture" from cross-sectional surveys.

The effectiveness of evaluations could be substantially increased with a greater awareness of information required by evaluations, and more care in designing surveys and selecting indicators to be monitored. With careful planning, this need not require a net increase in staff resources.

Ethical considerations

The objectives of saving lives and reducing suffering mean ethical considerations play an important role in design and implementation of emergency programmes and often override considerations which normally play an important role in the provision of development assistance. For instance, relief workers tend to regard cost-benefit considerations with distaste. Intensive feeding of severely malnourished children is very expensive per beneficiary and make substantial demands on the time of skilled personnel, possibly diverting staff and funds from programmes that would perhaps save the lives of much larger numbers of children. Nevertheless, the chance of providing assistance to children who are close to death is frequently regarded as sufficient justification for most relief programmes to include intensive feeding components. If decisions are made on the basis of non-economic factors, then evaluation techniques developed within an economic framework are, if not redundant, then hard to use.

Use of control groups is an important technique in analysing the impact of interventions. However, techniques such as experimental design and non-equivalent group design are ethically unacceptable. Selection of individuals or households to be deliberately excluded from a relief programme, thereby subjecting them to unnecessary suffering and possibly threatening their lives, so as to facilitate impact assessment, is unacceptable.

Relief programmes operate amid rapid change and considerable uncertainty. Randolph Kent has written that "the vast majority of disasters are marked by uncertainty, contradictory information and ambiguity... critical unknowns - affected populations, damage assessment, needs - are prone to serious distortions and contradictory evidence".

Setting objectives for an operation before it commences is hard. During a relief programme, individual agencies may alter key aspects of their programme as a result of changed circumstances which were not, or could not be, anticipated during the needs assessment/intervention design stage. For instance, a delay in the arrival of food aid for general use might lead an agency to establish a targeted supplement-

tary feeding programme in an attempt to prevent a deterioration in the nutritional status of children. Alternatively, an increase in the number of agencies operating in the affected area may, indeed ought to, lead to a rationalisation of activities geographically or in terms of type of activity.

In the face of uncertainty many agencies describe objectives only in very general terms, preventing evaluators testing performance against stated aims. In the rare instances where agencies set themselves specific objectives and identify the indicators to be used in assessing programme performance, the results are often unsatisfactory. For instance, CARE's 1992 relief programme in three provinces of northern Kordofan set specific and ambitious goals for its Supplementary Feeding Programme (SFP). One intermediate goal was "to increase the nutritional status of all children under five years of age registered in the SFP to exceed 85 per cent of weight for height before discharge from the programme." The final report on the programme notes: "There were no discharges from the programme as beneficiaries did not receive a consistent supply of supplementary food while irregular growth monitoring meant that data was generally inadequate to determine discharge eligibility."

Nature of institutional relationships

Many development projects involve formal agreements between only two or three agencies (donor, department of cen-

tral government and perhaps a local authority), a relief programme, even when operating in a limited area, typically involves a large number of organisations and agencies (donor, UN, government, one or more parts of the Movement and NGOs). Formal agreements usually exist "vertically" between donor organisations and implementing agencies. "Horizontal" relationships between organisations are often informal, and – depending on context, personalities, type of agencies – may include misunderstandings and rivalry. Lack of formal agreements, particularly when agencies are undertaking complementary components of an overall programme, reduces the "points of reference" available to evaluators to compare performance against expected roles and responsibilities.

The roles of the various organisations and agencies are often complementary. For instance, bilateral donors are involved in provision of funding and in kind resources, UN agencies are involved in assessment and coordination as well as resource provision and programme implementation, government agencies may be involved in needs assessment, coordination, resource provision and programme implementation, and NGOs in needs assessment, resource provision and programme implementation. There are two particular implications. First, effectiveness of coordination arrangements are an important area for evaluators to examine. Second, it is difficult to determine the relative contribution of different but complementary activities to such "out-

Box 5.2 Rwanda – assessing success and failure

The largest and most comprehensive evaluation of humanitarian response to a disaster is underway, aiming to assess the effectiveness of emergency assistance for the Rwanda emergency and look for lessons applicable to future complex disasters. The evaluation is costing less than 0.02 per cent of the value of the total assistance provided so far.

Unique in scale and scope, the evaluation was proposed by the Danish government assistance agency, DANIDA, and is backed by all major donor institutions. It will look at the work of all the humanitarian agencies involved throughout 1995, examining issues on a regional rather than country-by-country basis. The final report will be available in December 1995.

Explaining why a thorough evaluation of the humanitarian response to Rwanda's crisis is needed, DANIDA emphasised: "The human suffering is of an incomprehensible scale. The recent escalation of the

conflict and violence has received wide international concern and attention. The international community has provided substantial assistance to alleviate the human suffering and has contributed to efforts to find a peaceful solution to the conflict."

Donor assistance for Rwanda's emergency was expected to be well over US\$ 1 billion in 1994 or about 2 per cent of all Official Development Assistance (ODA). DANIDA saw two main reasons for the evaluation:

- Donors need to account for the relevance, effectiveness and impact of their aid,
- Valuable lessons may be learned for the planning and execution of future relief operations, as well as for ongoing operations in Rwanda and the region.

Separate studies will examine four areas and then be drawn together as one final report:

- Historic background, outlining the roots and cause of the conflict

within Rwandan society and seeking to identify possibilities for reconciliation;

- Emergency-prevention efforts, their effectiveness and relevance, including early warning systems and conflict management for monitoring and responding to approaching emergencies;

- Mechanisms for, and effectiveness of, preparation and coordination of emergency assistance programming and the impact of emergency assistance. This will concentrate on the effectiveness of coordinated action as well as how timely and appropriate the assistance proved as it was delivered through numerous channels,

- Planning and preparation for repatriation and rehabilitation to recreate and consolidate the capacities of those affected, reconstruct communities and launch sustainable development programmes to ensure more secure living standards than before the disaster. ■

come" indicators as the rates of morbidity, mortality and malnutrition.

In such a context there are obvious limitations to the value of evaluations which focus only on the activities of particular agencies. Major collective evaluations of collective relief efforts have, so far at least, not been undertaken, with the notable exception of the very recent commitment to a multi-donor evaluation of the Rwanda relief effort (see Box 5.2). The case for joint evaluations was clearly illustrated in southern Africa in 1993 and 1994 when at least a dozen agencies undertook independent evaluations of their response to the 1991/92 drought. A multi-agency overall evaluation with agency-specific case studies would probably have represented a more cost-effective and valuable approach for the international relief system as a whole.

An additional limitation on evaluations is the convention that only the UN evaluates itself and its specialised agencies. While assistance channelled through NGOs by bilateral donors is increasingly likely to be evaluated or audited, donors have been prevented from undertaking their own evaluations of assistance provided to or channelled through particular UN agencies

High political and media profile

Relief programmes in areas of conflict require donor agencies to consider factors such as sovereignty, international law, the "balance" of aid between opposing sides and perhaps also national foreign policy interests, factors which would not normally be considered when responding to a natural disaster in a peaceful country. Evaluations of such responses involve evaluators in examining matters normally be kept out of the public domain. Most relief programmes, whether in response to natural or conflict-generated disasters, have a high media profile which may also influence decisions on the timing, scale and nature of the response. Media profile and political issues make the evaluation process for relief programmes far more sensitive than evaluations of most development programmes.

A consequence is that many relief programme evaluations do not enter the public domain. While useful to the commissioning agency, it limits the evaluation's broader accountability function. If taxpayers (in the case of bilateral donor agencies) and private donors and the general public (for NGOs) are unaware of an evaluation's results, pressure on the agency to act on the findings are reduced. Evaluation reports are usually the property of the commissioning agency, even for the majority of "external" evaluations undertaken by "independent" specialists. Where evaluation reports do enter the public domain it may well be because they are considered to be a generally fa-

vourable assessment or because they are an abridged version of the original report with "sensitive" sections removed or rewritten.

How then should relief programmes be evaluated? A starting point has to be that evaluations are seen to be crucial to the process of learning from experience and improving upon future performance. The evaluation process and the information requirements for effective evaluation must be given higher priority than has been the case. If relief agencies are genuinely committed to improving their performance, then evaluations should be undertaken more frequently. One way might be for agencies to establish a ratio relating evaluations to the number or cost of relief operations undertaken, such as formally evaluating 50 per cent of relief operation spending.

Evaluation effectiveness could be substantially enhanced if the information needs of the evaluation process were approached in a more strategic way. Relief agencies and donor organisations should devise information strategies regarding the types of information required for programme monitoring and management, and for subsequent evaluations. Given multiple sources of funding, there should be agreement among donor organisations and relief agencies on appropriate indicators to monitor and the "basic minimum" information to be collected.

One route could be a working group of principal donor organisations and relief agencies to identify key indicators which all relief agencies should seek to monitor. Agencies might then incorporate the agreed set of indicators within their own Logical Framework Analysis methods. This would facilitate inter-agency comparisons and assessments of the overall effectiveness and impact of relief programmes. This in turn would help collaborative evaluations, cutting the cost per agency and providing agencies with a mechanism to jointly learn the lessons from relief programmes in which they are involved.

The prospects of evaluations being able to demonstrate the impact of relief programmes are limited. In most cases impact assessment will be qualitative. Agencies may have to recognise that the search for "proof" of positive impact of programmes involving many different components is unrewarding and concentrate their attention on more rewarding areas, such as giving more explicit statements of objectives in relief programmes so that outcomes may be tested. The highly dynamic context of most relief programmes will probably require agencies to prepare successive Logical Framework Approaches.

Another area where current practice could be improved is that of efforts by evaluators to seek the views of affected people. It will not usually be possible to undertake extensive surveys capable of

yielding statistically valid results. However, the views of a reasonably representative sample of people regarding appropriateness, effectiveness and timeliness of relief, the extent to which it supported coping and recovery mechanisms, and the effectiveness of any targeting methods may be gained simply by talking to a selection of beneficiaries in the project area.

A checklist would ensure that all households are asked the same questions but discussion around particular points should be encouraged. A broader group discussion on the relief programme may be helpful. In targeted assistance programmes, interviewing beneficiaries and non-beneficiaries could see how well they

conformed to the targeting. Such methods have provided rich insight into the effectiveness of particular programmes.

Finally, there remains considerable room for joint evaluations by bilateral donors, multilateral agencies and NGOs to objectively assess effectiveness. Though institutional barriers to joint evaluations are considerable, it is hard to defend the resources devoted to independently commissioned yet substantially overlapping evaluations of the same overall programme. Relief programmes are examples of the international community responding to a jointly perceived need and should be assessed in these terms. ■