

September 1998

RELIEF AND REHABILITATION NETWORK

7

GOOD PRACTICE **RRN**
review

**Evaluating
Humanitarian
Assistance Programmes
in Complex Emergencies**

by Alistair Hallam

Edited by
John Borton, Laura Gibbons and Sarah Longford



Comments should be sent to:

Relief and Rehabilitation Network
Overseas Development Institute
Portland House
Stag Place
London SW1E 5DP
United Kingdom
Tel: +44 (0) 171 393 1674/47/40
Fax: +44 (0) 171 393 1699
Email: rrn@odi.org.uk
Web site: www.oneworld.org/odi/rrn/

A copy will be sent to the author.

Comments received may be used in future Newsletters.

Subsequent versions of this Review will, at the Coordinator's discretion, take account of comments and suggestions received.

ISBN: 0-85003-373 X

Price per copy: £10.00 (excluding postage and packing)
(£7.50 to RRN members or for 10 or more copies)

© Overseas Development Institute, London, 1998

Photocopies of all or part of this publication may be made providing that the source is acknowledged. Requests for commercial reproduction of Network material should be directed to ODI as copyright holders. The Network Coordinator would appreciate receiving details of any use of this material in training, research or programme design, implementation or evaluation.

Abstract

Evaluation is a key tool in efforts to improve accountability and performance in the operation of the international humanitarian aid system. However, humanitarian aid has been subjected to less rigorous monitoring and evaluation procedures than those applied to development aid. As the share of overseas development assistance allocated to humanitarian aid has risen, and awareness of the complexity of humanitarian assistance has increased, so the need to develop appropriate methodologies for the evaluation of relief aid has become more apparent.

This Good Practice Review is the written output of an OECD/DAC project initiated to identify and disseminate best practice in the evaluation of humanitarian assistance programmes. The study seeks to improve the consistency and quality of evaluation methodologies, and enhance the accountability function of evaluation, contribute to institutionalising the lessons learned, and identify better methods for monitoring performance of humanitarian aid operations.

About the Author

Alistair Hallam is an economist with nine years overseas experience in Sudan, Somalia, Mozambique, Tanzania and Vietnam with UN agencies, local governments and NGOs, predominantly in relief management roles.

During 1995 he formed part of the core-team undertaking Study III "Humanitarian Aid and Effects" of the Joint Evaluation of Emergency Assistance to Rwanda. While preparing this present study, he worked as a Research Fellow within the Humanitarian Policy Group in ODI and provided support to the RRN team by taking editorial responsibility for a Network Paper, contributing articles to the Newsletter and assisting with the preparation of the Update, Publications and Conference sections. In addition he also participated in an evaluation on rehabilitation and war-to-peace transition issues in Angola and Mozambique undertaken by COWI Consult, the Chr. Michelsen Institute and the London School of Oriental and African Studies.

Alistair now works as an Economic Adviser at the UK Department for International Development (DFID).

Acknowledgements

Many thanks are due to the large number of individuals who participated in meetings where drafts of the Good Practice Review were discussed, and who sent comments to the author. The consultation process was extensive, and it is impossible to mention by name all those who helped. However, all contributions were very much appreciated.

A large vote of thanks must go to all my colleagues within the Humanitarian Policy Group of ODI. John Borton, Laura Gibbons, Caroline Dobbing, Sarah Geileskey and Sarah Longford have all played key roles in the production of this GPR, generously sharing with the author their expertise and their time. Their friendship has also been much valued. In addition, thanks to Helen Awan, Koenraad Van Brabant, Jo Macrae, Nick Leader, Aidan Cox and Angela O'Brien (all of ODI).

A debt of gratitude is owed to Niels Dabelstein who encouraged us at ODI to develop the proposal and who then guided the work through the DAC, in his role as Chair of the DAC Working Party on Aid Evaluation. His input into this work really began much earlier, when he was one of the key people involved in commissioning and managing the Joint Evaluation of Emergency Aid to Rwanda. Interest in 'good practice in evaluating emergency programmes' arose as a result of this exercise.

Two meetings were crucial in helping to shape the final output: a meeting in Copenhagen in January 1998, jointly organised by Danida and ODI, attended by members of the DAC Working Party on Aid Evaluation and by representatives of NGOs, the UN and the Red Cross movement; and a meeting in Canberra in March 1998, funded by the Australian Government and attended by donor representatives and a small group of people actively involved in humanitarian assistance evaluation.

Special thanks are due to those reviewers and commentators who showed a continued interest in the work throughout the process and who always had time to discuss ideas and offer advice: Andre Griekspoor (MSF-H), John Kirkby and Phil O'Keefe (ETC UK), Raymond Apthorpe (ANU), Jerry Adams (Tear Fund), Susan Purdin (SPHERE Project), John Telford (EMMA Ltd.), Claes Benedich (SIDA) and John Eriksson (independent).

Both Lowell Martin (UNHCR) and Des Gasper (ISS) kindly let me use parts of their work in the Review; Izzy Birch (Oxfam) provided useful case study material; Dr Steve Collins (independent), Bruce Laurence (MERLIN), Janice Giffen (Oxfam), Dr Nancy Mock (Tulane University), Richard Blewitt (British Red Cross), Claude Forthomme (FAO), Gary McGurk (SCF-UK), Susanne Jaspars (independent), Peter Hawkins (SCF-UK), Chris Leather (independent), Polly Byers (USAID), Kate Alley (UNICEF), Ted Kliest (Ministry of Foreign Affairs, Netherlands), Kari Karanko (Finnida), Dr Horst Breier (BWZ), Chris Raleigh (DFID), Ian Shaw (University of Cardiff), Antonio Donini (OCHA), Dr Marco Ferrari (Swiss Disaster Relief Unit), Peter Walker (IFRC), Jacques Stroun (ICRC), Sara Davidson (People in Aid), Hans Lundgren (OECD) and Louisa Chan (WHO) all provided valuable help at various stages. A large number of staff at FAO, WFP, the Emergency Unit of the Italian Ministry of Cooperation, MSF-F, ACF, Epicentre, the Cellule d'Urgence at the French Ministry of Foreign Affairs, LWF, ICRC, IFRC, OCHA, DFID, OXFAM, the World Bank, USAID, ActionAid, World Vision -UK, SCF-UK, CARE and the British Red Cross gave up time to be interviewed, or to respond to questionnaires.

Thanks are due to Danida, Finnida, DFID, The Netherlands Development Corporation and the RRN for funding the work that led to this GPR.

Despite the valuable assistance of all those mentioned, ultimate responsibility for errors and omissions rests solely with the author.

The Review is dedicated to Ann Gray for her support and patience.

Preface

This Good Practice Review (GPR) has been through an unusually long validation process. The process began with the circulation of a questionnaire in July 1997 to 50 key individuals involved in humanitarian aid evaluations and the analysis of the 16 responses (as well as a review of approximately 70 humanitarian assistance evaluation reports and synthesis studies).

This was followed by interviews with approximately 65 individuals in 30 (funding, channelling and implementing) humanitarian agencies in Paris, Rome, Geneva, New York, Washington and the UK. Preliminary findings and early

drafts of this report were presented at a number of fora involving evaluation managers and humanitarian organisations. Finally, successive draft reports have been commented upon by a range of individuals¹. The whole process has taken over one year. As a consequence, the findings do not simply represent the views of the author, but reflect a wider range of opinion.

The first substantive discussion of the project took place at the October 1997 meeting of the Active Learning Network on Accountability and Performance in Humanitarian Assistance (ALNAP) in London². This was followed by a meeting convened by Danida under the auspices of the DAC Working Party on Aid Evaluation in Copenhagen in January 1998, involving the representatives of 16 bilateral and multilateral donor organisations, 6 UN agencies, the IFRC and the ICRC and 5 NGOs. This meeting was particularly helpful in alerting the author to the different needs of different groups within the international humanitarian aid system and providing valuable feedback on aspects of good practice.

The study also benefited considerably from a workshop supported by AusAID and jointly organised by the National Centre for Development Studies at the Australian National University, Canberra and ODI which brought together a small group of evaluation specialists and practitioners with experience in evaluating complex emergencies for four days of discussions on the management of humanitarian aid evaluation. In addition the work was presented to the UK NGO Evaluation Group (REMAPP), the Evaluation Working Group of the UK Development Studies Association and to the April 1998 Annual Forum of the US NGO umbrella organisation, InterAction.

This GPR is just one of the products of an ODI-implemented project to identify and disseminate best practice in the evaluation of humanitarian assistance programmes. This project was approved by the Expert Group on Aid Evaluation of the OECD's Development Assistance Committee (DAC) in April 1997, in order to enhance the effectiveness of the evaluation process in the humanitarian field. It was funded by Danida, Finnida, the Netherlands Development Cooperation, the UK Department For International Development and ODI's own Relief and Rehabilitation Network.

Contents

	Key Findings and Recommendations	11
	The aim, purpose and outcome of an evaluation	11
	Planning an evaluation in the wider context	12
	The timing and scope of an Evaluation	12
	Implications for programme monitoring	14
	Communication – a vital two-way process	14
	Follow-up to an evaluation	16
1	Introduction	17
	1.1 Audience and scope	19
2	Benefits and costs of evaluations	23
	2.1 Benefits	23
	2.2 Costs	25
3	Key considerations	27
	3.1 Scope and nature of an evaluation	27
	3.2 The appropriate mix of policy and project evaluation	30
	3.3 Taking into account humanitarian space	33
	3.4 Assessing protection	34
4	Measures to enhance the effectiveness of the evaluation process	37
	4.1 Articulating and clarifying objectives	38
	4.2 Improving the availability of monitoring information	39

5	Establishing the basic parameters	43
5.1	What to evaluate	43
5.2	Accountability or lesson learning?	44
5.3	Evaluation: a peer review, not judgement from above	45
5.4	Clarifying the goals of an evaluation	45
5.5	Timing	46
5.6	Preparation, pre-studies and self-evaluations	47
6	The terms of reference	49
6.1	Preparing the terms of reference	49
6.2	Evaluation criteria	52
6.3	Making trade-offs explicit	57
7	Management of the evaluation	59
7.1	Management structure	59
7.2	Management process	60
8	Composition of the team	63
8.1	Composition of the team	63
8.2	Insider vs outsider	65
9	Methods of working	67
10	Sources of information and information-gathering techniques	71
10.1	Preliminary data collection	71
10.2	Interviews	72
10.3	Working in areas of conflict and instability	73
10.4	Gaining the beneficiaries perspective	74

11	Analysis techniques	79
11.1	Impact	79
11.2	Cost-effectiveness	85
11.3	Standards, norms and benchmarks	90
12	Writing the report, preparing recommendations and presenting findings	93
12.1	Writing the report	93
12.2	Structure of the evaluation report	95
13	Follow-up	97
	Annex I	99
	Annex II	105
	Annex III	109
	Acronyms	115
	Endnotes	116
	References	119
	Further Reading	123

Index of Figures and Boxes

Figure 1	The appropriate mix of policy and project evaluation techniques vary with the scope of the study	31
Figure 2	Resource flows within the international relief system	102
Figure 3	A standard logframe	106
Figure 4	A modified logframe	108

Boxes

1	Definition of Terms	18
2	Complex emergencies	21
3	System-wide evaluations	28
4	Discourse analysis	29
5	How a lack of monitoring information can limit the effectiveness of evaluations	41
6	Timing the evaluation to maximise its impact	47
7	The major share of humanitarian assistance is provided in conflicts	51
8	A lack of coherence between the humanitarian and military domains?	55
9	Joint consultations with stakeholders	72
10	Talking to all the stakeholders	73
11	Examples of different, and intensely-held perspectives	74
12	The gap between theory and practice	76
13	Talking to beneficiaries	77
14	Understanding local power dynamics	78
15	Differences in data availability	81
16	Disaggregating data	84
17	Counting all the costs (and all the benefits)	87
18	Examples of cost-effectiveness analysis	89
19	Code of conduct for the International Red Cross and Red Crescent Movements and NGOs in disaster relief	91
20	The SPHERE Project	92
21	Following-up evaluations	98

Key Findings and Recommendations

The aim, purpose and outcome of an evaluation

- Evaluations offer a chance to reflect upon the events leading to a humanitarian crisis and the responses to that crisis, and to determine whether procedures can be altered to improve responses in the future.
- Formal evaluations result in written reports, which contribute to enhancing transparency and accountability, and allow for lessons to be learned across programmes and agencies.
- Given the high turnover of staff in emergency situations, evaluations provide a means for organisations to retain and build institutional memory.
- Evaluations are a way of examining the effectiveness of relationships within an organisation. They provide staff an opportunity to discuss issues of concern without prejudicing their position.
- Evaluations, if correctly carried out, with the beneficiaries' perspective as a central component of the process, can go some way towards improving the system's accountability 'downwards'.
- Evaluations of humanitarian assistance programmes should be seen as a contribution to a dialogue about programme performance, and not as a "judgement from above".

Planning an evaluation in the wider context

- The capacity to carry out “wider-scope”, system-wide evaluations, in a regular and organised manner needs to be developed within the international humanitarian system. System-wide evaluations are particularly important where local governments are weak, where coordinating structures for the international response do not exist or lack support, or for regional emergencies where no regional coordination structure exists. Individual project evaluations can fail to pick up on “system-wide” features, such as the overlap of activities, gaps in provision, and the lack of a global strategic framework.

- Humanitarian agencies’ willingness to move from a position of offering relief on the margins of a conflict, to becoming considerably more involved in the theatre of conflict itself, requires evaluators to acknowledge the wider political context of emergencies, and the potential political impact of humanitarian programmes.

- It is important for the terms of reference to be grounded in the broader questions concerning the nature of the problem and the effectiveness with which it was addressed. The way in which those questions specific to each evaluation are framed is likely to evolve, as the study proceeds. The terms of reference should, therefore, be treated as negotiable by the Evaluation Managers and the Evaluation Team.

The timing and scope of an evaluation

- Clarity over evaluation objectives will make the whole process clearer and easier. Evaluation Managers should allow adequate time for this critical stage of identifying the objectives of the study.

- There are strong arguments in favour of carrying out an evaluation of a humanitarian assistance operation or a complex emergency while it is still ongoing rather than waiting for the operation to wind down.

- Humanitarian assistance is essentially a 'top down' process. Humanitarian agencies are often poor at consulting or involving members of the affected population and beneficiaries or their assistance. Consequently, there can often be considerable discrepancy between the agency's perception of its performance and the perceptions of the affected population and beneficiaries. Experience shows that interviews with beneficiaries can be one of the richest sources of information in evaluations of humanitarian assistance. Interviews with a sample of the affected population should be a mandatory part of any humanitarian assistance evaluation.

- Evaluations should comment on the impact of humanitarian aid programmes, and not focus solely on monitoring the efficiency of project implementation. However, humanitarian programmes often take place in complex, confused circumstances, where plans change regularly, where information is scarce and where little is predictable. As a result, qualitative and deductive methods of measuring impact, that involve beneficiaries, are likely, in general, to be more appropriate than methods that seek to be 'scientifically rigorous'.

- Evaluations of humanitarian assistance programmes should include an assessment of measures to provide protection to the affected population. In many emergency situations, the first needs of a population under threat may be protection from murder and harassment, as well as from discrimination that can lead to exclusion from basic services. The evaluation should also assess what measures have been taken to mitigate potential negative consequences of the humanitarian programme.

- Evaluations of humanitarian assistance programmes can benefit from the use of the additional evaluation 'sub-criteria' of 'connectedness', 'coverage', 'coherence', 'cost-effectiveness', and 'appropriateness'.

Implications for programme monitoring

- Monitoring systems for humanitarian assistance programmes should take full account of the needs of evaluations. This will require:
 - i) the use of data collection systems that facilitate evaluations and cross-agency, cross-programme comparisons;
 - ii) agreement to be reached on the indicators that should be monitored by all agencies. Potentially the Sphere Minimum Technical Standards project under way within the NGO community will provide such agreement on key indicators;
 - iii) a commitment by implementing agencies to facilitate evaluations through the management of filing and information systems so that key reports showing the decision-making process are easily accessible.

Communication – a vital two-way process

- The emergence of key issues throughout the evaluation requires flexibility of approach, with good lines of communication between the Evaluation Manager and the team on an on-going basis.
- Evaluators need to be aware of how difficult it can be for those caught up in conflict and instability to discuss what they have experienced. Regardless of who the team is interviewing, it is important that they bear in mind, and as far as possible empathise with, the experiences that their interviewees have endured.

- In evaluations of humanitarian assistance programmes, interviews are generally the most important source of information on what happened and why. Evaluators need to talk to a wide range of 'stakeholders', to build up as complete and balanced picture of an intervention as possible: agency staff, local NGO partners, host government officials – national and provincial – and UN officials.

- At the end of the evaluators field visit, there should be a workshop where the main findings are presented to the stakeholders.

- For accountability purposes, and for reasons of clarity, evaluators should make sure that the link between their findings and the evidence used is clear. Output data should be presented, whenever available, not only for accountability purposes, but also to allow the reader to make his or her own judgement as to the overall findings of the evaluator. The logic of the evaluator should be clear and convincing. At the same time, it must not be assumed that evaluators are infallible – methodological clarity enhances the accountability of the evaluator to those being evaluated.

- A draft report should always be prepared and circulated for comment to those organisations and individuals involved in the evaluation. Adequate time is needed for it to be considered and for comments to be received.

- Evaluations of humanitarian assistance programmes should include an assessment of the links between costs and performance. Cost-efficiency analysis is generally not possible for 'relief' programmes even where 'full' cost-effectiveness analysis is not. Too many evaluations make recommendations in the absence of any sense of the costs associated with alternative approaches.

Follow-up to an evaluation

- To maximise the effectiveness of the evaluation process, there needs to be an organised follow-up, that tracks responses to recommendations made in the evaluation report.

- Evaluations are of potential interest to agency managers, beneficiary representatives, local governments, donor officials, students of humanitarian issues, the press, as well as the general public. Ideally, each audience should receive a different products tailored to their needs. While the availability of resources for publication will generally limit what can be done in this respect, it is important to underline that attention to the dissemination of evaluation results is as important as carrying-out the evaluation in the first place.

- Evaluation reports need to be “sold”. Donors, field officers and agency staff need to be enthused, excited and convinced that the evaluation report is important and should be read. While selling the report is more the responsibility of the management group than the evaluation team, marketing strategies could be included in negotiated follow-up actions in order to help steering committee members sell the evaluation report within their own organization.

1

Introduction

The OECD/DAC³ define evaluation as “an assessment, as systematic and objective as possible, of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, developmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors” (see Box 1, overleaf).

When looking at the *design* of the programme, the evaluator needs to assess how the problem was identified and analysed by the donor and/or implementing agency: who was consulted, what skills did they have, what was the information base on which decisions were taken? How was it decided which methods of intervention, partners and procedures should be used? How were priorities set? What assumptions were made, and on what basis? Were assumptions linked to risk analysis?

Assessment of the *implementation* process involves looking at staff – their skills and the way that they are organised and operate. Were decision-makers

Box 1

Definition of Terms⁴

- Aid is **efficient** if 'it uses the least costly resources necessary to achieve its objectives'. This implies, *inter alia*, that 'the aid can gain the most results for its economic contributions'.
- **Effectiveness** of aid relates to the effects of aid vis-à-vis the objectives set. Aid is effective to the extent that the objectives are achieved.
- **Impact** is a wider term. It refers to the 'effect (of an aid intervention) on its surroundings in terms of technical, economic, socio-cultural, institutional and environmental factors'.
- **Sustainability** refers to the extent to which the objectives of an activity will continue (to be reached) after the project assistance is over'.

able to access important information? Were the right skills available to allow actors to analyse information and respond appropriately? Logistics, procurement and strategy – and decision-making processes will need to be assessed: were outputs achieved in the most efficient manner? How was the relationship between donors and implementing partners? And between partners in the field? Did actors share information on a regular basis? It will also include assessing the monitoring, reporting and evaluation procedures and the way an agency interacts with beneficiaries.

Finally, the **results** need to be examined. Were programme outputs appropriate? Was the programme effective? What was the impact of the intervention on the lives of beneficiaries? Did the programme achieve what it set out to do?

In an ideal world, the evaluator would find that there was adequate documentation and institutional memory to allow for consideration of all these elements of a programme, that programme objectives were clearly stated from the outset, that indicators that could measure the achievement of objectives were selected in advance, and that a monitoring system was established from the outset that resulted in the collection of the appropriate data. In practice,

however, none of these conditions may apply. While there are measures that can be taken to improve the effectiveness of the evaluation process (see Chapter 4 on page 37), the evaluation of humanitarian assistance programmes, particularly those undertaken in an environment of conflict and instability, will always present significant challenges. It is these challenges that occupy the bulk of this Good Practice Review (GPR).

1.1 Audience and scope

This GPR is aimed at those involved in the evaluation of humanitarian assistance programmes, whether in NGOs, official aid departments, the UN system or Red Cross organisations. Inevitably, not all of the material in the GPR will be relevant to all those in the intended audience. However, the humanitarian system involves many actors working interdependently and it is impossible to target one specific audience alone. This means that recommendations have been made that are beyond the capacity of one individual organisation, and whose fulfilment may require co-operation amongst a number of actors.

The report is intended both to encourage the practice of evaluation and to make the evaluation experience a productive one. The GPR does not attempt to describe in detail all the possible evaluation tools discussed. There are numerous texts that provide guidance on specific evaluation issues (such as participatory evaluation techniques, for example), and it is recommended that the reader wishing to know should consult such texts. A list of further reading can be found on page 123.

This GPR is designed both for those undertaking evaluations, and for those working on programmes being evaluated. The author of this review has been involved in a number of evaluations of humanitarian assistance programmes and has found that there is often considerable uncertainty, amongst field-based personnel, as to the rationale and justification for the evaluation. On occasions, this is manifested in terms of some hostility towards the evaluators. More often, however, it means that the evaluation process is not as effective as it could be, and that field-based staff lose an opportunity to use the evaluation process to raise important issues and bring about needed change in programmes.

This GPR focuses on the evaluation of humanitarian assistance programmes rather than on development aid programmes more generally. While there is much in common between development and humanitarian aid evaluations, there

are also sufficient differences to justify a publication dedicated to the latter, particularly in light of the substantial increases in expenditures on humanitarian assistance over the last decade. The GPR also concentrates primarily on ‘complex emergencies’ (see Box 2, opposite).

This GPR is designed to help programme managers who are concerned with both improving performance and increasing accountability in respect of their programmes. It is motivated by the experience of a number of evaluators that has shown that humanitarian assistance evaluations are often limited in impact because: they are ‘added-on’ at the end of a programme, rather than being built into programme design (also common for evaluations of development aid); key data on which to build analysis is lacking; key staff have moved on by the time the evaluation is carried out; key questions are not asked; they are unnecessarily rushed, with inadequate time to consult with those most affected by programmes, and to allow for adequate feedback once the report is written; and, they fail to comment substantively upon impact.

Box 2

Complex emergencies

The term 'complex emergency' was coined in Mozambique in the late 1980's as a way for aid agencies to acknowledge that humanitarian assistance needs were being generated by armed conflict as well as by periodic 'natural disaster' events, such as cyclones and droughts, while avoiding the use of terms such as 'civil war' and 'conflict' which were sensitive terms in the Mozambican context at the time. Since then, the term 'complex emergency' has entered widespread usage as a way of differentiating those situations where armed conflict and political instability are the principal causes of humanitarian needs from those where natural hazards are the principal cause of such needs.

While some consider the use of the term 'complex emergency' unhelpful, as it implies that some disasters are not 'complex', the term is useful in that highlights the fact that situations involving political instability and armed conflict do make a difference to the way in which humanitarian problems can be tackled. These differences, in turn, have implications for the evaluation process.

A 'typical' complex emergency is characterised by:

- a collapse of state functions
- intra-state rather than inter-state conflict
- difficulty in differentiating between combatants and civilians
- violence directed towards civilians and civil structures
- fluidity of the situation on the ground
- a lack or absence of normal accountability mechanisms
- the potential and actual development of war economies
- the potential for humanitarian assistance to prolong the conflict
- a multiplicity of actors

See Annex II on page 105 for further discussion.

2

Benefits and costs of evaluations

2.1 Benefits

The OECD/DAC guidelines on aid evaluation⁵ state that “the main purposes of evaluation are: to improve future aid policy, programmes and projects through feedback of lessons learned; to provide a basis for accountability, including the provision of information to the public”.

Evaluations are an integral part of the project cycle, and a key means by which agencies seek to learn lessons from their work and then incorporate these into future policy and practice. Such ‘organisational learning’ is a pre-requisite for more wide-reaching ‘institutional learning’, a process by which lessons are learnt between and across agencies⁶. Given the high turnover of staff in emergency situations, evaluations provide a means for retaining and building institutional memory. The rapid growth in expenditures, combined with policy uncertainty of how to respond effectively to complex emergencies, has created a particular need to invest in ‘learning from experience’ at this stage in the development of the humanitarian assistance system.

Humanitarian agency workers, particularly in long-running emergency situations, can find themselves working in a relative state of isolation. The

extent of their work and their achievements can be known only to very few outside their immediate group. This can lead to a situation where fundamental questions and assumptions go untested, with the result that the impact of the work may be reduced. Evaluations offer a chance to reflect upon the events that lead to a humanitarian crisis and the responses to that crisis. Where people are involved in making important decisions in the face of what can be significant pressures, it is important that there is an opportunity to pause to review whether these decisions were the right ones, and whether procedures can be changed to encourage better decision-making in the future.

It is particularly important for the international (humanitarian) system collectively to create the space in which to learn lessons from events that may have seemed beyond its control at the time. For example, when a million Rwandans entered Goma in July 1994, agencies simply had to put all their energies into the task of providing assistance. The question of why the influx happened, and whether the response could have been improved still needed analysis, albeit something only possible some time after the events in question.

Evaluations serve as an important means of documenting what actually happened in an operation. Indeed, they are often the only consolidated source that provides an historical account of how a programme progressed. The fact that formal evaluations result in written reports contributes to enhancing transparency and accountability, even on those occasions where the evaluation reveals little that was not already known by the staff involved. In addition, a written report means that lessons can be shared more easily across programmes and agencies, and come to form valuable case study examples for training courses and teaching programmes.

The link between those closest to a problem, on the ground, and those, perhaps in headquarters, who are taking key decisions, is a critical one. If this breaks down, there can be serious consequences for programme effectiveness. Formal evaluations are an important way of assessing this link, to check whether messages from the field, whether from agency staff or beneficiaries, are working themselves upwards to policy makers. Problems can arise because of poor information systems, or because those at field level are cautious of making strong criticisms of programmes that may reflect upon their colleagues, immediate superiors or those who may control the resources they need for survival. An evaluation should offer field staff the space to discuss issues of

concern to them which can then be presented more ‘neutrally’ by the evaluator. Where ‘outsiders’ are involved, new perspectives can be gained on old problems.

Humanitarian assistance tends to involve a very ‘top-down’ approach, with beneficiaries often barely involved in management decisions. Accountability often tends to be ‘upwards’ to donors and senior managers, with very little ‘downward’ accountability to beneficiaries. Evaluations, if correctly carried out, with the beneficiaries’ perspective as a central component of the process, can go some way towards improving the system’s accountability ‘downwards’. Such issues are discussed at greater length later in the paper.

2.2 Costs

On the downside, evaluations are often seen as imposed from above, costly and unnecessary. There is a fear, among humanitarian agency workers, that an external evaluation will result in judgements being made about their work by individuals who do not understand the complexity of the situation. There is also a concern that, even where the evaluation makes a correct assessment, it will not make any difference, and will be filed and forgotten.

There is no easy response to such criticisms. There is no denying that some evaluations are poorly prepared and carried out, absorb valuable staff time in the field, and result in poor quality reports. However, there is also no doubt that other evaluations – even when imposed on an agency by donors – can be of considerable benefit. By improving the way in which evaluations are carried out, and by enhancing field staff’s understanding of the evaluation process, it is hoped that this GPR will increase the proportion of evaluations providing real benefits to programmes. At the same time, there is a need for the organisations involved to make efforts to improve the way in which they learn, so that lessons emerging through the evaluation process feed through into improved policy and practice.

It needs pointing out – to those who object to the spending of aid resources on evaluations – that the costs of *not* regularly assessing programmes can be very high, in financial and human terms. Cost-effectiveness analysis during the Rwanda evaluation (Borton et al, 1996), for example, revealed that freight costs from Indian Ocean ports to the refugee areas were US \$70 -85 per tonne using rail wherever possible, compared to US \$115-34 per tonne by road. Given that approximately 170,000 tonnes of food were transported in 1994

alone, the impact on budgets of using the right channel were enormous. The OLS Review (Karim et al, 1996) found that reductions in emergency food aid designed to increase self-reliance were not justified on the basis of information gathered, and were having a possibly serious negative effect on beneficiaries. The evaluation led to important changes in the programme.

It is difficult to be categorical about the amount of money needed to conduct an evaluation. One consultancy company that has carried out a number of evaluations recommends, as a rule of thumb, that 2% of total programme costs should be assigned for evaluation (with 4% for monitoring)⁷. One of the most expensive humanitarian assistance evaluations of recent times – the Joint Evaluation of Emergency Assistance to Rwanda⁸ – cost over US \$1 million. However, this was not even one tenth of one percent of the volume of funds spent in the Great Lakes in the first year of the emergency response. Even if the result is only small improvements in the humanitarian system, benefits can significantly outweigh costs.

Evaluations are also sometimes used as a substitute for good management, with evaluators finding that they are expected to gather data and make recommendations on issues that should have been resolved much earlier. Evaluations can also be used cynically to ‘rubber stamp’ decisions already taken. Neither use is appropriate, and is only likely to leave most of the participants dissatisfied and frustrated with the evaluation process.

3

Key considerations

3.1 Scope and nature of an evaluation⁹

There is wide variation in the scope of humanitarian assistance evaluations, with evaluations of a single project undertaken by a single agency at one extreme and total system evaluations of the response by the international community to the whole conflict at the other. In the middle are single bilateral programme studies and partial system studies, as well as thematic evaluations (e.g. programmes for refugees, or IDPs, etc.) and sectoral evaluations (health interventions, etc.).

Audit, evaluation and review

A distinction is commonly made between ‘audit’, ‘evaluation’ and ‘review’: unlike reviews, evaluations involve a more thorough assessment of impact, the writing of a report, and the involvement of individuals not normally associated with the day-to-day running of the programme. Reviews are generally more regular, less onerous, and sometimes purely verbal processes. Both are important components of a management system, and complement each other. Traditionally audits tend to be associated with financial accountability, and concentrate more upon individual honesty and integrity, rather than the broader achievements of a programme¹⁰.

Box 3**System-wide evaluations**

It is important to note that there is currently no capacity in the system for regularly promoting and organising 'system-wide' evaluations, despite recognition of the importance of system-wide factors in influencing the effectiveness of a humanitarian response. The Joint Evaluation of Emergency Assistance to Rwanda, for example, was carried out only due to the efforts of a few key individuals, who pushed the process forwards.

However, the nature of the findings of the Rwanda evaluation underline the importance of system-wide evaluations, particularly where local governments are weak, where coordinating structures for the international response do not exist or lack support, or for regional emergencies where no regional coordination structure exists. The performance of individual agencies may be good, in terms of meeting their own objectives, but major problems may arise with regard to the overlap of activities, gaps in provision, and the lack of a global strategic framework. Individual project evaluations can fail to pick up on these 'system-wide' features. 'Wider-scope' evaluations are needed that look outside the present institutional frameworks used for managing humanitarian assistance programmes, as these frameworks are often a reflection of historical processes, rather than the most effective structure in which to provide assistance.

The logistics of organising a large-scale multi-agency evaluation, however, can be considerable, and there is arguably only a finite number of such studies that the system as a whole can absorb. Such evaluations do not, in any case, substitute for programme evaluations. Perhaps the lesson for managers in the field, is to consider how their evaluation plans might be harmonised with the evaluation plans of other agencies operating in the same area, in order that the process is more effective. Coordinating evaluations also has the benefit of limiting duplication (particularly where a programme has been funded by a number of donors who might otherwise all wish to conduct their own evaluations), thus cutting costs and saving staff time.

Policy or project evaluations

Another important distinction commonly made between types of evaluation is whether they are ‘policy’ or ‘project’ evaluations. While such a distinction is somewhat artificial, it is useful as a means of highlighting the range of techniques required of evaluators.

Policy evaluation, broadly speaking, focuses on principle-, rationale- and objective-setting, and predictive and explanatory theory. It looks at the framework of understanding, beliefs and assumptions that make individual projects possible as well as desirable. The focus of policy evaluations often includes the mandates of agencies and their ideologies and institutions, often in a framework which compares one agency or set of agencies with another. Project evaluation, on the other hand, focuses predominantly on the actual performance of operations set-up within such a framework (indeed, often assuming that the framework is essentially unproblematic).

Policy evaluations seek out the inherent tensions or contradictions in policy objectives, through tools such as discourse analysis and logic-of-argument analysis. For example, a donor may seek to combine relief, development and peace objectives in its policy towards complex emergencies (see Box 4 below). Discursive, philosophical, theoretical analysis may show these to be counteractive or contradictory. At the same time, if empirical evidence demonstrates that actual operations have been successful, despite convincing theoretical argument against a policy being justifiable, then policy evaluation should take this into account and policy be modified accordingly.

<p>Box 4</p> <p style="text-align: center;">Discourse analysis</p> <p>Discourse analysis involves the recognition of the fact that ‘there is a plurality of values and arguments available for thinking about any specific policy issue. Analysis, therefore, has to be part of a process in which these several points of view are taken into account or directly included in the analysis.’ [White,1994 in Gasper and Apthorpe, 1996]. Discourse analysis involves a high level of multi-disciplinary analysis, in comparison to more traditional research methods.</p>

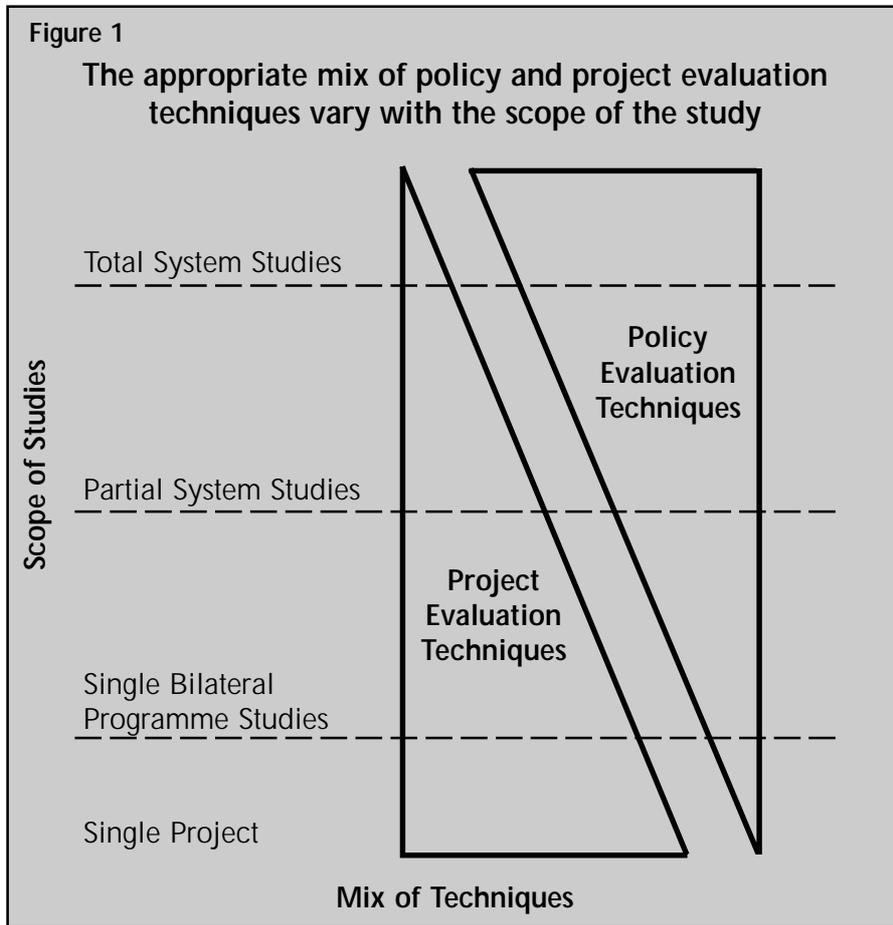
Policy evaluation involves a process of ‘validating’, through argument, rather than ‘verifying’, through some ‘scientific’ process, the various interpretations. In this respect, it is a ‘discipline comparable with judicial procedures of legal interpretation, having a logic of uncertainty and of qualitative probability’ (Ricoeur, 1979:90 in G&A). Policy evaluation, being concerned with the sum that is qualitatively greater than the (project) parts, is generally less handicapped by the lack of quantitative data than project evaluation.

Policy goals or objectives are seldom clear or sharp enough to serve as criteria against which to test management and performance. They are not immediately evaluable in their own terms. In particular, those involved in policy evaluation must take the stated objectives of policies, projects and evaluation ToR, and translate them into something evaluable. Recognising that goals and objectives may change in response to changing circumstances on the ground is another requirement for best evaluation practice. This is particularly true of complex emergencies, which are, by their nature, unstable, yet also very often protracted or recurring.

3.2 The appropriate mix of policy and project evaluation

In the past, evaluations of humanitarian assistance tended to focus on projects rather than on policy issues. Donor organisations were generally more comfortable evaluating the projects through which a policy was implemented, rather than the assumptions which lay behind a particular policy. However, several recent studies have explicitly considered policy issues (e.g. Apthorpe et al 1996; Karim et al 1996). Given the range of policy questions currently confronting those involved in the provision of humanitarian assistance (e.g. how best to provide protection to civilians in an ongoing conflict; how best to move from relief to development; whether humanitarian assistance does actually prolong conflicts), there are strong arguments for more direct consideration of policy issues in the evaluation process.

All evaluations will generally involve a mix of policy and project evaluation techniques. However, (see Figure 1, opposite) the precise balance between the two will depend upon the scope of the evaluation. Total system studies will invariably involve use of policy evaluation techniques, though with the use of detailed project-level analysis to illustrate the points being made. At the same



time, single project studies cannot simply ignore the context in which operations take place. The evaluation of humanitarian assistance *programmes* – the most common form of evaluation – involves a degree of both project and policy evaluation techniques.

The humanitarian response system is a complex one, composed of numerous interdependent relationships. Explanations based on the separation of cause from effect are often not possible because the direction of influence is often circular rather than linear. Thus methods which are more common in historical or philosophical research are often more productive than those traditionally employed in the social sciences. Such methods acknowledge the complexity

and interdependent nature of events in the real world and ask not, ‘did ‘x’ cause ‘y’?’ but rather, ‘what happened?’ and ‘why?’. Attempts are made to understand situations and structures, to analyse a particular set of events and processes by constructing a narrative history to serve as a kind of model or spectrum through which to view what is to be understood and explained. A range of different actors may be asked to tell their story, recognising that what they say represents the truth as they see it, or the truth as they would like the evaluator to see it. In this way, a partial understanding of someone else’s view of reality may be developed. The stories of many different actors are then added to the available documentary evidence to construct the narrative history. The narrative history is therefore more than a simple chronology. It details not just what happened and when, but also who was involved, and why, and links significant events.

The narrative history is used by the evaluator to help to make judgements about events and processes, to explain why actors did what they did and with what effects and draw practical conclusions accordingly. When many different stories are accumulated, consistent patterns may emerge which help to explain actions. Alternatively, a framework for understanding events may arise out of previous studies and be confirmed by the evidence revealed in the narrative history.

The skills required for policy evaluators includes skills in discourse and argument analysis, and the ability to generalise ‘lessons learned’ from one situation to another – that is, in comparative and deductive analysis, much as expert witnesses testify when serving public and judicial enquiry, drawing on general principle and norms established elsewhere and applying them to the specific case.

The narrative history approach described above is particularly suited to evaluating complex emergencies because it focuses on qualitative as well as quantitative methods; because it allows the evaluator to focus on actions, processes and intentions; and because it highlights the competing agendas of the diverse range of actors. Furthermore, use of the narrative history approach does not preclude the use of more analytical models, such as logical frameworks or cost-effectiveness reviews, to examine specific components of an emergency assistance programme.

3.3 Taking into account humanitarian space

Humanitarian aid programmes are subject to many constraining influences, for example: governments may block resources going to rebel areas; transport infrastructure may be poor; rains may close roads for months at a time; information on population movements may be difficult to obtain. Any evaluation of performance needs to take into account these constraints that limit the humanitarian space available to the agency.

Humanitarian space is a dynamic concept. Levels of access and availability of resources can change regularly during an aid operation. Agencies can themselves influence the humanitarian space available to them. Successful negotiation, for example, may open new routes through rebel areas. Agencies can also reduce the space available to them by sticking rigidly to their mandate, even though flexibility might prove more effective. Agencies may sometimes take a principled stand and refuse to supply relief inputs where an unacceptably high proportion of these are being diverted by combatants to fuel the war. Humanitarian space may thus be restricted in the short-term, in the hope that this will lead to more freedom to operate effectively over the longer-term.

There are likely to be considerable institutional and political constraints that reduce the ability of relief agencies to achieve the most (cost-)effective solutions. For example, in late 94-early 95, Kibumba refugee camp in Goma received its water supplies through an extremely expensive tanker operation, despite the fact that a hydrological survey had shown that it would have been cheaper to invest several million dollars in constructing a pipeline to the camp from a perennial water source several kilometres away or, alternatively, to move the refugees closer to Lake Kivu. However, for political reasons, the Zairean authorities did not want to signal to the local population that the refugees were going to be staying for an indefinite period and were therefore unwilling, regardless of the cost (borne largely by the international community), to bring them closer to the lake or encourage the building of 'permanent' water systems.

3.4 Assessing protection

Evaluations of humanitarian assistance programmes should include an assessment of measures to provide protection to the affected population. In many emergency situations, the first needs of a population under threat may be protection from murder and harassment, as well as from discrimination that can lead to exclusion from basic services. The evaluation should also assess what measures have been taken to mitigate potential negative consequences of the humanitarian programme.

The need for protection may be independent of the humanitarian assistance programme. For example, a humanitarian assistance programme may be ongoing in an area where there is a campaign of ethnic cleansing but bear no relation to this fact. An evaluator should still assess, however, whether agencies are aware of the context in which they are operating, and what they are doing to enhance protection. This may involve lobbying for international action, providing witness to atrocities or informing local authorities of their obligations under the Geneva Conventions. Not every agency needs to be involved in such explicit protection activities however. It is perfectly reasonable for there to be a division of labour among agencies – with perhaps the ICRC taking the lead in issues relating to contravention of International Humanitarian Law. The task of the evaluator is to assess whether due consideration has been given to the issue of protection.

In some emergencies, the need for protection is more directly related to the humanitarian assistance programme. For example, the perpetrators of the genocide in Rwanda in 1994 were able to locate some of their victims by observing where relief deliveries were taking place. Several years later, in late 1996, those fleeing the refugee camps in eastern Zaire (now Democratic Republic of Congo) came out of their hiding places in the forests to receive food and medical treatment from the aid agencies. Many were then killed. In both cases, the victims were killed not because they were receiving aid, but because the aid programme helped the perpetrators of these crimes to locate their enemies. In other instances, the act of receiving aid has made populations vulnerable to harassment or even death. In Liberia, a food delivery in a contested area was followed by a massacre of the population who had received the food. The aim was not simply to steal the food, but to demonstrate to the population that, although they might be able to receive aid from the international community, they were still under the control of the armed group concerned.

An evaluation should assess an agency's analysis of the security problem and its approach to dealing with it, just as it should assess an agency's response to other significant problems.

4

Measures to enhance the effectiveness of the evaluation process

A theme of many recent humanitarian assistance evaluations has been that the effectiveness of the evaluation process has been limited, in many cases significantly, by the lack of :

- i) clearly stated objectives for the overall programme and its various sectoral and project components, and
- ii) monitoring information necessary for assessing the performance of projects and thus of the overall programme.

As a result, there is often insufficient basis on which to judge an intervention, a lack of data with which to measure impact, and vague or non-existent objectives against which to measure outcomes – all of which reduce the extent to which evaluations can meaningfully comment on impact and performance.

It is important to emphasise, however, that improving the articulation of objectives and the availability of monitoring information is essential in improving programme performance in general. They should certainly not be seen as something to be done solely in order to improve the evaluation process. Evaluations are not a substitute for effective programme management, and should not be used as such.

4.1 Articulating and clarifying objectives

Encouraging explicit and clear statements of objectives will require concerted effort over a period of time at several different levels of the humanitarian system. The following planning tools can assist the process:

- **Strategic frameworks**, which set the overall objectives for the international community’s response. (A Strategic Framework is currently being developed and piloted by the UN system for Afghanistan);
- **Country or response strategies**, articulated by donor organisations and other organisations involved in the response to a particular country or complex emergency; and
- **Logical Framework Analysis¹¹**; LFA or ‘LogFrames’ are increasingly being used by donor organisations and some implementing agencies as a way of articulating the goal, purpose, outputs and indicators for humanitarian assistance projects. Some donor organisations have made the use of “LogFrames” mandatory for funding proposals above a specified funding level.

Use of the LogFrame does appear to: significantly increase transparency in the setting of objectives; make more explicit the conceptual frameworks underlying interventions; help to avoid confusion between means and ends; and highlight the ‘level’ at which an organisation is intervening. However, concerns have been voiced that, for humanitarian assistance programmes, the log-frame is over-restrictive, potentially inflexible and that while useful for looking at inputs and outputs, it has proved less useful in looking at overall policy goals, social ideals and higher principles. A number of modifications have been proposed to improve the usefulness of LFA for humanitarian assistance programmes. Briefly, these require greater attention to the assumptions/risks statements, which need revisiting as the project progresses, as well as during the evaluation process. These are discussed further in Annex II (see page 105).

4.2 Improving the availability of monitoring information

A product of the characteristics of complex emergencies is that key information on a range of matters of vital significance to evaluators is often unavailable. The multiplicity of actors, the fluidity of the situation, the difficulties of working in the context of war and instability, the frequent absence of baseline data on the condition of the affected/target populations prior to the assistance intervention, the failure of some agencies to monitor key indicators, a lack of agreement on standardised monitoring procedures and protocols among agencies, and the difficulties of adhering to normal standards for recording discussions and decisions and maintaining filing systems, all combine to produce a situation where information which is vital to evaluators either does not exist or is not easily accessible. Evaluators of humanitarian assistance programmes are, as a consequence, faced not only with the need to compensate for missing project data, but also to contend with a lack of information on the context, the precise sequence of events during the period and the goals and (often changing) policies of different actors at different stages of the emergency.

A number of the difficulties listed above are due simply to the nature of the context in which humanitarian assistance programmes are undertaken. As discussed earlier, this requires evaluators to adopt a variety of instruments, and to construct narrative histories and ‘pictures’ of the vital information to serve as a form of baseline from which to judge the appropriateness and effectiveness of policies and projects. However, there are also measures which can be taken to enhance the effectiveness of the evaluation process, including more attention to the monitoring process.

Monitoring and evaluation are distinct exercises. Monitoring systems should meet the on-going information needs of programme managers, and should not be seen as existing to provide information solely for evaluation purposes. This said, however, in a context of rapid changes in the environment in which a humanitarian programme is being undertaken, it is important that programme managers regularly collect information that allows for judgements to be made about the overall appropriateness of the programme. Current monitoring and reporting systems for humanitarian assistance programmes often do not do this. Indeed, too often monitoring systems do not provide the information required for routine management decisions to be made. There are a number of

reasons for such weaknesses: a lack of time in the early stages of a response to either design an adequate monitoring system or to collect data; a lack of emphasis on the need to collect data; and, a lack of understanding by field staff on what data to collect.

Addressing these problems will require concerted action throughout the humanitarian system to ensure that:

- i) implementing agencies improve their monitoring systems and use data collection systems that facilitate ex-post evaluations and cross-agency, cross-programme comparisons;
- ii) agreement is reached on the key indicators that should be monitored by all agencies. Potentially the SPHERE Minimum Technical Standards project – see section 11.3 on page 90 and Box 20 on page 92 – under way within the NGO community will provide such agreement on key indicators;
- iii) a commitment by implementing agencies to facilitate evaluations through the management of filing and information systems so that key reports showing the decision-making process are easily accessible.

Too often, evaluators start their work unaware of all the relevant internal documentation or useful related studies undertaken by other agencies in the same area or similar agencies in the same country. Sometimes, the problem is simply one of poor file management. For instance it is not uncommon for evaluators in the field offices of operational agencies to be shown a room full of files in boxes, and told to sort out the good from the bad themselves. Valuable time can be spent searching for key documents.

It is essential, for the organisation’s institutional memory as well as for the evaluation process, that important information is retained and readily accessible. Evaluation Managers could seek to avoid such problems by insisting that agencies involved in the evaluation collate their files and information to be used by the evaluators well in advance of work starting. Key documents, such as Situation Reports, and Monthly Project Reports, should be located and filed where they can easily be accessed. Country offices involved in the evaluation should be included in such efforts as they are often well placed to know of related studies by other agencies.

Box 5

How a lack of monitoring information can limit the effectiveness of evaluations

In response to the 1991-92 drought in Mozambique, the UK ODA provided over 15 grants to NGOs for seeds and tool distributions. In examining narrative reports submitted to ODA the evaluation team found that only one of the agencies unambiguously indicated the dates when the seeds were distributed. It was, therefore, not possible for the team to state what proportion of the total tonnage of seeds had reached farmers before the onset of the rains. Only one of the agencies supported undertook surveys during the harvesting period to assess the production levels achieved and thus the overall effectiveness of the seeds and tools interventions. Thus two key indicators of effectiveness and impact were not available for the majority of the agencies supported.

The dysentery epidemic which affected the Great Lakes Region during 1994 was responsible for more deaths than any other single non-violent cause. Mortality rates in the IDP camps in the *Zone Turquoise* appear to have been very high. However, many of the camps were not covered by epidemiological surveillance systems until after the epidemic had peaked. For instance surveillance in Kibeho, the largest and most problematic camp, only started 11 weeks after the area became secure. As a result, the precise dimensions and dynamics of the dysentery epidemic in this important camp of around 100,000 people will never be known. Given this lack of information it was not possible for evaluators to assess the effectiveness and impact of the nutritional, medical and public health interventions carried out by the agencies which worked in Kibeho and the other camps.

5

Establishing the basic parameters

5.1 What to evaluate

Potentially there are many options available to Evaluation Managers in deciding the scope, focus and objectives of an evaluation. Should only the response to the emergency be considered or should the actions that were, or might have been, taken prior to the emergency also be considered? If the emergency was prolonged over several years should the evaluation look back over the whole emergency or just consider the most recent or critical periods of the emergency?

There is a tendency for implementing agencies to evaluate only those projects they were involved in implementing and for donor organisations to evaluate only those projects which they supported through the provision of financial, material and logistical assistance. However, many other possible approaches exist and should be considered. The highly inter-related nature of the different types of intervention produces strong arguments in favour of collaborative studies with other agencies, enabling related interventions to be considered together. For instance, it is rarely possible to assess the impact of food aid on morbidity and mortality without also taking into account health and sanitation programmes. Similarly, it rarely makes sense to consider the effects of a programme in one village without taking into account programmes in

neighbouring areas, to which those villagers may have had access. However, collaborative studies are organisationally more demanding and require a degree of consensus and trust on the part of the agencies. Furthermore, even for larger-scale evaluations, not all components of a response can be considered. In such cases, the evaluation should focus on the worst bottlenecks in the system, as a way of highlighting issues with potentially important structural implications.

While there can be no hard and fast rules on which projects or programmes to select for evaluation, a number of considerations should be borne in mind:

- all major humanitarian aid programmes should be evaluated;
- innovative projects, or projects that have generated considerable comment, whether critical or enthusiastic, should be evaluated.

5.2 Accountability or lesson learning?

Another question to be considered is whether the evaluation should emphasize lesson-learning or accountability, or a mix of the two. In evaluation literature, it is often argued that the evaluation process, the projects selected and the composition of the evaluation team will all vary depending upon the primary purpose of the evaluation – lesson-learning or accountability: where the primary purpose is accountability, evaluators will need to be independent, committed to objectivity, and focused on results and outputs; where the purpose is to learn lessons that can be fed back into the management process with the aim of improving future performance, evaluations will be more participatory, with the evaluation team including members of the programme staff, and the focus more on the programme process rather than the outputs.

Cracknell (1996), suggests that where evaluations are for accountability purposes, there should be a random sampling of projects selected for analysis. A fair cross-sectional representation of projects is required, and the work should be carried out by totally independent evaluators from outside the organisation. For lesson-learning purposes, Cracknell argues for the deliberate selection of projects with problems, or deemed to be of particular interest, the work involving greater use of own staff.

In practice, however, many evaluations seek to fulfill both lesson-learning and accountability objectives. Indeed, a number of evaluators of humanitarian

assistance programmes do not find the distinction to be particularly relevant: ‘accounting’, without a willingness to learn lessons is not very useful. At the same time, you can generally only draw lessons after you have given account of what you have done and why. Furthermore, in order to improve an ongoing programme, it is necessary to look at impact as well as process. Given the difficulties of ‘measuring’ impact in any sort of ‘objective’ way, many evaluations end up gathering what information they can on outcomes and then look at process and try to deduce what parts of the process have led to what outputs. This inevitably involves close work with programme staff, even a certain degree of ‘negotiation’ over what is likely to have been the impact of various actions taken. The distinction between ‘participatory’ and ‘objective’ evaluations becomes blurred.

5.3 Evaluation: a peer review, not a judgement from above

Whether evaluations of humanitarian assistance programmes are carried out for accountability or lesson-learning purposes, they should be seen as a contribution to a dialogue about programme performance, and not as a ‘judgement from above’. Evaluations are not ‘audits’ in the traditional sense of that word (concerned with adherence to rules, often focusing on financial accountability). An evaluation’s recommendations do not generally concern the breaking of a recognised rule, and often require further debate and consideration by those involved.

5.4 Clarifying the goals of an evaluation

Clarity over the objectives of the evaluation will make the whole process clearer and easier. Evaluation Managers should allow adequate time for this critical stage of identifying the objectives of the study. Advice from a range of sources inside and outside the organisation should be drawn upon. Evaluation Managers with little previous experience of evaluating humanitarian assistance programmes ought to involve a humanitarian assistance evaluation specialist early on in the process or consult with other Evaluation Managers with more experience. However, the process of agreeing the objectives will have to be carefully managed to avoid a ‘shopping list’ of unprioritised and potentially conflicting objectives.

5.5 Timing

When an evaluation is best carried out will depend in large part on the objectives of the study and the context of the operations to be studied. However, there are strong arguments in favour of carrying out an evaluation of a humanitarian assistance operation or a complex emergency while it is still ongoing rather than await the winding down of the operation and the end of the emergency: the need for evaluators to construct a narrative and ‘baseline’ requires that they interview many of the key actors before they are redeployed to operations in other parts of the world (the turnover of humanitarian agency personnel being notoriously high); the ability of the target population and beneficiaries of assistance to accurately recall events such as the timing and routing of their displacement, the timing of their first relief distributions and the identity of the providing agency is reduced with each week that passes after the peak of the operations. Ideally, therefore, the first field visits should take place within 6 months of the peak of the operations, though the elapsed times for many of the humanitarian evaluations undertaken to date is between 12-18 months. Such early deployment will probably involve careful consideration of the security risks and insurance considerations, as well as greater attention to sensitive questioning where interviewees may have been through traumatic events. However, the quality of the evaluation is likely to be greater if it is not delayed significantly.

Even where conditions may not be ideal for an early evaluation, there is a lot to be said for a reasonably ‘quick and dirty’ evaluation, to take place, while programmes can still be changed. Evaluations are not carried out solely to improve on-going programmes, of course, but also to learn more general lessons about policy and implementation, for future interventions. However, an evaluation that can help re-orient an on-going programme has added significance (see Box 6, opposite). Field workers may also feel more willing to engage with the evaluation if they feel that the benefits are more immediate.

Evaluations require a certain amount of time to tender, assemble teams, refine terms of reference and make travel arrangements. If they are to happen within 12 months of a response, then planning will need to start shortly after the programme itself starts.

It may seem trivial, but it is important to be aware of other, sometimes seasonal events in the country where the field visit takes place: in the rainy season, it

Box 6

Timing the evaluation to maximise its impact

OXFAM UK conducted a participatory evaluation of their programme in Ikafe, northern Uganda, in mid-1996, some eighteen months after the programme began. The review prompted OXFAM to reconsider the programme's underlying assumptions, and led to the development of more achievable aims and objectives. It also brought to the fore some tensions which needed addressing, particularly with local communities. The timing of the review meant that there was sufficient programme experience upon which to draw, but also sufficient time subsequently to put lessons into practice. The participative way in which it was conducted, with a follow-up workshop four months later, also meant that the views and concerns of a wide range of stakeholders could be heard and accommodated.

Source: Neefjes, K. and David, R. (Oct 1996), 'A participatory review of the Ikafe refugee programme', OXFAM, Oxford.

may not be possible to visit certain areas; there may also be a time of the year when government and/or agency staff traditionally take their holidays, and will be out of the country.

5.6 Preparation, pre-studies and self-evaluations

Key informants, namely those working on programmes about to be evaluated, agencies or donors associated with the programme, as well as local and national government officials and representatives of the affected population, should be warned in advance of the arrival of an evaluation team. This will increase the possibility of their availability and give them the chance to prepare relevant documents, and reflect on their experiences.

The complexity of the subject matter and the need to construct narratives and baselines calls for flexible and tailored approaches to the humanitarian assistance evaluations. One possible approach is to stagger the evaluation so as to enable a first phase or 'pre-study' to construct the narratives and baselines and identify the key issues on which to focus in the main phase of the study. Experience has shown that some of the main issues that emerge in humanitarian assistance

evaluations are not apparent at the start of the process. Because of the need to interview key agency personnel and begin collecting key documentation, pre-studies will probably involve visits to the site of operations.

Another approach towards identifying the key issues is to request personnel involved in the operation to carry out some form of self-evaluation process. This approach has the merit of making the personnel being evaluated feel included and valued in the process. This can help them incorporate lessons arising from the evaluation. Such an approach should not prevent the evaluation team from focusing on issues not identified through the process of self-evaluation.

Such an iterative approach requires flexibility on the part of Evaluation Managers. The Terms of Reference may need to be modified or renegotiated during the process. Good quality and frequent communications will need to be established between an Evaluation Manager and the evaluation team.

6

The terms of reference

6.1 Preparing the terms of reference

While it may often be tempting to prepare ‘minimalist’ terms of reference (ToR) which simply reproduce the standard evaluative criteria of effectiveness, efficiency and impact, it is important for the ToR to be grounded in the broader questions concerning the nature of the problem and the effectiveness with which it was addressed. However, the way in which those questions specific to each evaluation are framed is likely to evolve, as the study proceeds. It is important, therefore, that the ToR should be treated as negotiable by the Evaluation Managers and the Evaluation Team.

Time needs to be set aside to amend the ToR so that all the stakeholders are satisfied with what is to be attempted. This may require some flexibility in funding arrangements. Alternatively, the use of the staged approach or a pre-study ‘scoping’ phase indicated in section 5.6 (on page 47) may be more appropriate. The ToR for the main phase of the evaluation may be finalised in the light of the findings of the pre-study.

The issue of who should be involved in drawing up the ToR is particularly difficult. Broad ownership of the evaluation can be generated if the development of the ToR is an inclusive process. However, different actors frequently have

different agendas and these agendas will inevitably influence the ToR. How far these different agendas should be accommodated is by no means fixed. In some situations, it may simply not be possible to reconcile differing agendas.

Evaluators (and those commissioning evaluations) will have to realise that there may not be consensus on what should be in the ToR, and debates around this issue may reflect debates about the programme itself. Particular areas of contention may need highlighting early on, so that evaluators can be on the look-out for evidence that either supports or refutes a particular view (Wood, 1998). It is important for evaluators and evaluation managers to be realistic about what is possible from an evaluation. It may be better to spend more time on fewer issues, than attempt to cover a huge agenda.

Given the context of conflict and insecurity, protection issues are critical to the effectiveness of humanitarian action. Where levels of protection are poor, it is feasible that the target population of an otherwise effective relief project is being killed by armed elements operating in the project area or even within the displaced persons/refugee camp. Assessment of the level of security and protection in the project or programme area and, where relevant, the steps taken to improve them, should be part of all humanitarian assistance evaluations. In the humanitarian assistance evaluations undertaken to date, such issues have often been left out or inadequately covered. Often the lack of familiarity of Evaluation Managers with the issues of security and protection has contributed to such omissions.

International agreements on standards and performance (see section 11.3 on page 90), such as the Red Cross/ NGO Code of Conduct (see Box 19 on page 91), the People In Aid Code of Conduct in the Management and Support of Relief Workers¹² and the Sphere Project (see Box 20 on page 92), as well as relevant aspects on international humanitarian law provide international norms against which the performance of agencies and the system may be assessed. The ToR ought, therefore, to refer the evaluators to such norms.

The ToR should be realistic: they can become wish lists, where far too much is demanded of evaluators given only a short time in the field. Unrealistic demands can lead to a situation where the team ignore the ToR, and just focus on what they feel are the most important aspects of the programme. (Lindahl, March, 1998) This can lead to dissatisfaction with the results and disillusionment with the evaluation process. At the same time, it is important to be aware that limiting

the focus can also cause problems. A recent evaluation of MSF operations in Kivu decided to focus on operations rather than on advocacy, as this had already been investigated previously. However, removing such an important feature of the programme from the picture gave a distorted impression of how resources had been used, as resource use was analysed only in respect of project outputs (personal communication with Andre Griekspoor, MSF-Holland). In Annex III, a checklist of questions commonly asked in evaluations is presented (drawing heavily from Martin, L. 'Planning and Organising Useful Evaluations', UNHCR, January 1998). The checklist is useful as a way of stimulating thought, and suggesting potentially new areas of inquiry.

The political dimensions of humanitarian aid

Humanitarian agencies' willingness to move from a position of offering relief on the margins of a conflict, to becoming considerably more involved in the theatre of conflict itself (see Box 7 below), requires evaluators to acknowledge the wider political context of emergencies, and the potential political impact of humanitarian programmes. For example, supplying aid to those in a besieged town may enable them to resist for longer than would otherwise have been the case, and allow one side in a conflict to gain military advantage over another. Supplying assistance to the refugee camps of Goma, in the absence of efforts by the Government of Zaire and the international community to separate the former Rwandan Government army and militia from the refugees allowed the former to regroup and renew attacks on Rwandan citizens. Relief assistance supplied to IDPs in Mozambique allowed the Government to maintain a presence in areas of the country where otherwise it would have had to withdraw.

Box 7

The major share of humanitarian assistance is provided in conflicts

The World Food Programme reports that in 1995, it spent roughly \$19 million worldwide on meeting the effects of natural disasters; this compares with more than \$1.5 billion on conflict-related emergencies (Webb, 1996). In 1993, the European Community Humanitarian Office, the world's largest humanitarian aid donor, spent over 95 per cent of its aid responding to conflict-related emergencies. (Randel, 1994).

Sometimes, it is not the physical aid itself which has the unintended political impact, but the mechanisms of providing aid that are exploited for political purposes. For example, to deliver aid behind rebel lines may involve negotiating with rebel leaders, thereby conferring a certain legitimacy upon them. In addition, rebels and government officials can use the fact that aid is being provided as a way of demonstrating their power to local population groups, without directly expropriating the aid itself.

As well as the effects outlined above, there are political effects of distributing to one group within a village and not another, of the channels used to provide assistance, of the choice of interlocutors within beneficiary populations. Sometimes aid can reinforce existing power structures. Other times it can challenge them. In both cases, poorly thought through aid can result in negative as well as positive impacts. Sometimes the negative impacts can work against long-term solutions to humanitarian problems.

6.2 Evaluation criteria

Evaluation criteria help frame the types of questions asked of a project or programme. Those used in the OECD/DAC definition of an evaluation – “an examination, as systematic and objective as possible, of an on-going 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” – have become the standard criteria for evaluating development programmes.

Efficiency

Measures the outputs – qualitative and quantitative – in relation to the inputs. This generally requires comparing alternative approaches aimed at achieving the same outputs, to see whether the most efficient process has been used.

Effectiveness

Measures the extent to which the project or programme achieves progress toward its purpose, and whether this can be expected to happen on the basis of the outputs of the project.

Impact

Looks at the wider effects of the project – social, economic, technical,

environmental – on individuals, communities, and institutions. Impacts can be immediate and long-range, intended and unintended, positive and negative, macro (sector) and micro (household). Impact studies address the question: what real difference has the project made to the beneficiaries? How many have been affected?

Relevance

Is concerned with assessing whether the project is in line with local needs and priorities (as well as with donor policy).

Sustainability

Of particular importance for development projects – sustainability is concerned with measuring whether an activity or an impact is likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable.

Whilst these OECD/DAC evaluation criteria are broadly appropriate for humanitarian assistance programmes, their usefulness may be enhanced by the addition of ‘sub-criteria’. [When this was proposed in earlier drafts, it generated considerable controversy. Some argued that there was no need to alter the DAC criteria, as these could be interpreted to incorporate the ‘sub-criteria’. Others, however, found that the new ‘sub-criteria’ helped them to focus on issues of particular relevance to humanitarian assistance programmes. Rather than be prescriptive, and come down one way or another, it seems sensible to present the sub-criteria below, and leave it to the reader to decide which criteria or ‘sub-criteria’, they feel are most appropriate. The Joint Evaluation of Emergency Assistance to Rwanda (Steering Committee of the Joint Evaluation, 1996) was one of the first evaluations to use the new criteria. Indeed, the DAC criteria were not explicitly used on this evaluation, and were replaced by the criteria of ‘connectedness’, ‘coverage’, ‘coherence’, ‘cost-effectiveness’, ‘impact’ and ‘appropriateness. An on-going DFID evaluation of the response to the volcanic eruptions in Montserrat in 1997 also uses the new ‘sub-criteria’. These are discussed below.]

Connectedness

This refers to the need “to assure that activities of a short-term emergency nature are carried out in a context which takes longer-term and interconnected problems into account” (Minear, 1994). Many humanitarian assistance

interventions, in contrast to development projects, are not designed to be sustainable. They still need assessing, however, in regard to whether, in responding to acute and immediate needs, they take the longer-term into account. For example, otherwise efficient food distribution programmes can damage roads used by local traders, while the presence of large refugee camps can result in severe environmental impacts in neighbouring areas. Local institutions can also suffer – the high salaries paid by international NGOs can attract skilled staff away from government clinics and schools, leaving the local population with reduced levels of service. Large-scale relief programmes can also have a significant impact on local power structures, both for better and for worse.

Coverage

This refers to the need “to reach major population groups facing life-threatening suffering wherever they are, providing them with assistance and protection proportionate to their need and devoid of extraneous political agendas” (Minear, 1994). Coverage alerts evaluators that complex emergencies and associated humanitarian programmes can have significantly differing impacts on different population sub-groups, whether these are defined in terms of ethnicity, gender, socio-economic status, occupation, location (urban/rural or inside/outside of a country affected by conflict) or family circumstance (e.g. single mother, orphan). Programmes need to be assessed both in terms of which groups are included in a programme, and the differential impact on those included. For example, studies have shown that, in Ethiopia in the 1980s, more than 90% of international relief went to government-controlled areas, penalising those in areas of Tigray and Eritrea controlled by insurgent movements (Minear, 1994). Other studies have revealed that single mothers may be disadvantaged when it comes to access to resources, as they are unable to leave children to queue for relief goods. In the case of the Great Lakes emergency, it was found that the coverage of the response varied enormously: refugees and IDPs, and residents in neighbouring IDP camps, were often treated in quite different ways, despite having very similar needs.

Coherence

This refers to the need for policy coherence, the need to assess political, diplomatic, economic and military policies as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian considerations.

The example in Box 8 (below) refers to a notable lack of coherence in the international community’s response to the Great Lakes emergency in 1994. In other instances, donor-imposed trade conditions have been blamed for precipitating economic crisis and conflict, undermining longer-term development policies. Coherence can also be analysed solely within the humanitarian sphere – to see whether all the actors are working towards the same basic goals. For example, there have been instances of one major UN agency promoting the return of refugees to their host country while another is diametrically opposed to such policies.

In addition to these sub-criteria, there are others issues that Evaluation Managers might consider making explicit within the ToR, such as the issue of *timeliness*. While timeliness is really implicit within the criteria of effectiveness, (for if the delivery of food assistance is significantly delayed the nutritional status of the target population will decline), there is value in using it more explicitly as one of the stated criteria because of its importance in the consideration of emergency programmes. Similarly, issues of resourcing, preparedness and coordination have been recurrent themes of humanitarian assistance evaluations undertaken to date and could be included in the ToR as potential areas of focus.

The Joint Evaluation of Emergency Assistance to Rwanda replaced the criteria of relevance with the criteria of *appropriateness* – the need “to tailor humanitarian activities to local needs, increasing ownership, accountability, and cost-effectiveness accordingly” (Minear, 1994). However, the two criteria complement rather than substitute each other. ‘Relevance’ refers to the overall goal and purpose of a programme, whereas ‘appropriateness’ is more focused on the activities and inputs. The addition of the new criteria drew attention to

Box 8

A lack of coherence between the humanitarian and military domains?

During the crisis in the Great Lakes region in 1994, military contingents were withdrawn from Rwanda during the genocide. Evidence suggests that a rapid deployment of troops could have prevented many of the killings and the subsequent refugee influx into Zaire. A huge relief operation then followed, once most of the killing was over.

the fact that even where the overall programme goal is relevant – for example, to improve nutritional status – there are still questions to be asked about the programme purpose. Distributing large quantities of food aid may not be the best way of improving nutritional status. Alternatives could include food for work, cash for work, or measures to improve the functioning of local markets. Furthermore, even if distribution of food aid is deemed appropriate, it is still necessary to examine the appropriateness of the food that is distributed. Is it culturally acceptable? Does it provide a balanced nutritional intake? Is it whole maize or ground maize (and are there milling opportunities for those receiving whole maize)?

It is also important to assess whether agencies take into account the safety of their own staff, not only from violence, but also from sickness and disease. For example, staff should not be expected to work for long periods in highly volatile areas – where their personal safety may be in doubt – without adequate support. Similarly, in a malaria-endemic area, staff should have information and protection against the disease. Where programmes involve spraying against vectors, staff should have the appropriate safety gear. At the same time, evaluations may need to consider the conduct and behaviour of local and expatriate staff, in particular in regard to cultural sensitivities.

Finally, there is the important issue of *coordination*. Given the multiplicity of actors involved in an emergency response, it is important that coordination is explicitly considered – the intervention of a single agency cannot be evaluated in isolation from what others are doing, particularly as what may be appropriate from the point of view of a single actor, may not be appropriate from the point of view of the system as a whole. For example, however well an organisation undertakes a curative health programme, impact is likely to be considerably reduced if water and sanitation concerns have not also been addressed. Similarly, while it may be appropriate for an individual donor organisation to focus on the most critical region in a country in crisis, it would be inappropriate for all donor organisations to do the same.

6.3 Making trade-offs explicit

In any programme there will be a number of trade-offs that have to be made between meeting different goals. For example, there may be a trade-off between connectedness and impact: providing large quantities of food aid may result in a greater impact in reducing food security, but raise dependence and cause damage to local markets. Similarly, there may be a trade-off between efficiency and appropriateness: taking food to a conflict area by road rather than by plane may be considerably cheaper, even if some of the food is stolen en route, but could involve negotiating and conferring legitimacy on rebel movements.

The task of the evaluator is to make these trade-offs explicit, highlighting the issues of concern and commenting on the decisions taken. This is important for accountability purposes as well as for lesson learning.

Management of the evaluation

7.1 Management structure

There are a number of ways of organising the management of evaluations. In some organisations, a separate and/or central evaluation unit is responsible for carrying out evaluations, whereas in others, responsibility for evaluation is decentralised throughout the organisation. The rationale for a dedicated unit is usually to ensure transparency and the independence of evaluations. In some cases, the head of the evaluation unit answers directly to the board, so that there can be no pressure put on this individual by senior management. In the most extreme cases, the heads of evaluation units are allowed only to serve a five-year term, after which they have to leave the organisation.

The argument against the model of having one central unit for all evaluations is that, while it may encourage high-quality and independent evaluations, it does not further institutional learning, as the responsibility for evaluations is taken away from the department responsible for implementing the programme and undertaking any reforms suggested by the evaluators.

There are a number of possible management structures of multi-donor or collaborative evaluations. For the Joint Evaluation of Emergency Assistance to Rwanda, a Steering Committee was formed, which included many of the

principal stakeholders – donors, UN and Red Cross agencies, and NGO umbrella groups. The Committee met approximately once every six months to review the process at key moments. The day-to-day management of the study was, however, delegated to a smaller management committee, comprising five donor evaluation officials. To protect the independence of the study (particularly important, given the political sensitivities concerned), the evaluators were required to *consider* all comments made on the drafts, but were not required to incorporate every comment.

Where evaluations are to include participatory work with beneficiaries, the involvement of local stakeholders in the management committee should be considered. For certain evaluations, it may be useful to involve government officials¹³. Again, there needs to be an understanding that involvement in a management committee does not signify a right to alter the contents of a report, merely to provide input at key stages. Particular care may be needed where ongoing conflict means that biases are likely to be too strong.

7.2 Management process

Approach

The more iterative emergence of key issues requires flexibility of approach, with good lines of communication between the evaluation manager and the team, on an on-going basis. Evaluation managers may need to devote comparatively more time to management issues for evaluations of humanitarian assistance programmes in complex emergencies: for discussions with the evaluation team; to act as a link between the team and the agencies (or emergency desk) involved; and, to negotiate changes in the focus of the terms of reference.

Time frame

Adequate time must be allowed for the entire evaluation process. Time is needed to prepare the ToR, to carry-out a preliminary review of available documentation, to commission pre-studies, to arrange and carry-out the field work, to write up the draft report, to circulate the draft for comments, to finalise the report and, ideally, to hold a workshop to discuss the findings. There is understandable (budgetary) pressure to ensure that these tasks are carried out in the shortest time possible, but this must not be at the cost of the evaluation itself. The time allocated to writing-up, following the field-trip, is often found

to be too short. It is important that time is allowed to send messages to the field to test emerging hypotheses and to fill in data gaps. Furthermore, it is important to ensure that all the team members have seen and discussed the report before it is circulated widely.

Preparation

The nature of the subject matter means that evaluations of the humanitarian aid responses to complex emergencies are undertaken in areas that have recently experienced active conflict and may be continuing to experience instability and insecurity. While it is rare for evaluation teams to be deployed to areas of active conflict, this is not unknown, and may require the provision of additional insurance cover for war zones to personnel not covered by schemes already available for the personnel of the agency commissioning the evaluation.

8

Composition of the team

8.1 Composition of the team

The composition of an evaluation team will obviously depend on the focus of the evaluation, and the nature of the programme being evaluated. In addition to ‘technical’ skills in specific relief sectors, such as medicine, or water and sanitation, there is a need for general evaluation skills, such as an analytical mind, knowledge of data-gathering procedures and report-writing skills. In addition, even the most ‘technical’ programmes, need to be evaluated for their wider political and social impact. For example, the quality of water produced is just one of the factors determining the impact of a water programme. Many other questions also need to be asked: how far do people have to walk to obtain water? Does everyone have equal access to water points? Do more powerful groups control access, or even charge for access? Are women safe from harassment when collecting water? Does the water programme fit in with an appropriate sanitation programme? (Beyond a certain investment in water production, resources spent on health education may have a greater impact on mortality and morbidity than extra resources spent on new water points.) Are requirements only for drinking water or also for showers, wash basins, etc.? Should user fees be introduced? Would it be better to rehabilitate and extend a town water system rather than investing in a new system for a temporary IDP camp?

The complexity of such issues, and the specialist knowledge required to evaluate different types of intervention and aspects of the programme implies the need for a fairly wide range of skills and specialist knowledge. But for evaluations of limited scope, focussing on just one type of intervention, a small team may be adequate. Where a large range of skills are needed, there are essentially two options available to Evaluation Managers. One is to assemble a large team and accept the management difficulties that are likely to flow from this. Another is to form a small core team of able generalists, and bring in sectoral expertise where necessary. This might be done by establishing an Advisory Group for the core team.

The gender balance of teams should be a consideration for evaluation managers. Ideally, teams should have a balanced gender composition. Evaluations must involve the beneficiaries, and this will generally mean considerable time will be spent with women (and children). While not always the case, it may be more effective, or culturally acceptable, for women beneficiaries to be interviewed by a female evaluator.

Regardless of the exact structure of the evaluation team, all team members should get together at several stages of the evaluation process to discuss overlapping issues and conclusions, for the strength of multi-disciplinary teams lies in the differing perspectives that can be brought to bear on the issues. Provision for such meetings should, therefore, be included in initial budget estimates.

Experience suggests that a tender process alone seldom produces the optimum team. Good team leaders are more likely to be produced by a process of short-listing suitable individuals, discussing their strengths and weaknesses, and then asking individuals if they are interested, or confining a bidding process to those on the short-list. In other words, a search process should be combined with a tender process.

Where larger teams are involved, team leaders are needed who have good general management skills as well as strong evaluation backgrounds. Experience of evaluating, or operating in, emergency programmes is also highly desirable. Team leaders also need to be confident communicators, able to manage the team, participate in negotiations over terms of reference, identify key issues for evaluators and be responsible for interviewing senior level key informants who can provide an overview of an emergency and its response.

There is a clear need for (at least, some of the) evaluators to have knowledge of the country concerned. However, when selecting a team, consideration must be given to whether the evaluators were ever previously involved in the programmes being evaluated. A number of major evaluations have encountered significant problems as a result of one of the team's prior involvement in the programmes under consideration. This does not mean that such people should not be included in evaluation teams, but they should not lead such teams. Country specialists can become stakeholders in aid programmes and sometimes function as apologists, or be overly defensive of dysfunctional aspects of humanitarian assistance in a particular country (Lindahl, March 1998).

The writing abilities of the Team Leader are also important. For instance, an evaluation of operations in Somalia by one bilateral donor ran into difficulties because the technical experts could not write up their sectoral reports and integrate them. Additional resources were then needed in order to complete the task. Instead of using the team leader as editor, an alternative option may be to involve a professional content editor early in the process.

There are benefits in including locally-recruited personnel in the evaluation team. Specific technical skills may be available from local institutions or NGOs, and such personnel should have the advantage of appropriate language skills, as well as an understanding of local culture and politics. The use of personnel from local NGOs and institutions can allow for valuable cross-project and cross-agency learning, as well as helping to build-up local evaluation capacity. It is important, however, whether using local or international evaluators, that there is an awareness of local ethnic and political sensitivities.

8.2 Insider vs outsider

Increasingly, evaluations are becoming more participatory and involve staff members of the organisation being evaluated. There are a number of advantages to such an approach: the experience of being part of an evaluation team can help NGO personnel develop their analytical and management skills; the use of programme staff can increase the legitimacy of what may otherwise be seen as a one-sided, donor-imposed process, which can limit the learning effect and result in the influence of the evaluation on the ground being close to zero. However, there are also occasions when people involved in a project will be too 'close' to the work to be able to be objective, or where those being questioned

will feel reluctant to talk critically of someone they know as a colleague, or superior or even ‘benefactor’.

Outsiders (whether completely independent or merely from a different part of the organisation), able to stand-back from day-to-day management concerns, can take a more dispassionate view of a project’s successes and failures. In addition, they provide stakeholders an opportunity to make confidential comments on the actions and attitudes of project managers. Even for staff or beneficiaries who do feel comfortable making criticisms outside of evaluations, the involvement of outsiders may increase the chance that senior policy makers will be able to ‘hear’ the opinions of junior staff, without these being filtered through numerous layers of an organisation. Evaluations, involving an input of extra personnel with different skills and experiences, provide an opportunity to focus on issues in more depth than usual, particularly where the routine demands of day-to-day management issues mean that regular staff are always busy.

In general, credibility is enhanced by the involvement of independent evaluators. It is important, therefore, for emergency programmes in contexts where other accountability mechanisms may be lacking (no free press or functioning judiciary, for example) that the team is, at the very least, led by an independent outsider. While it may appear most pragmatic to use mixed teams, a recent evaluation by a major donor encountered severe problems when the critical findings of the independent consultant on the team were edited out of the final report by the donor team leader. It is, however, important to be aware that ‘independence’ is a relative concept. Some consultants are well aware that a highly critical report about a donor or agency will not be well-received and may not lead to future contracts. This may affect their willingness to be entirely objective.

9

Methods of working

This section does not try to give a detailed description of the many evaluation methods and techniques available. There are a number of books that provide such guidance: Gosling, and Edwards, ‘Toolkits: A Practical Guide to Assessment, Monitoring, Review and Evaluation’; Kumar and Casley, ‘Project Monitoring and Evaluation in Agriculture’ and ‘The Collection, Analysis and Use of Monitoring and Evaluation Data’, Marsden and Oakley, ‘Evaluating Social Development Projects’, etc. (see references and additional reading on page 119). This section tries to give an overall framework for evaluation, in which the various evaluation tools and techniques would be applied. Readers are strongly encouraged to refer to the above books to supplement their reading of this review.

Preparing the narrative history and ‘baseline’ has to be the starting point for any study. Expansion and modification of the narrative history and ‘baseline’ will probably continue throughout the study as more information is obtained. While documentation will be an important source of information, interviews with the range of actors and members of the affected population will be a vital source too. Arguably, interviews form a more important source than is normally the case with evaluations of development assistance due to the problems of poor record keeping and documentation noted earlier. Effective management

of the results of these interviews is an important determinant of the effectiveness of the team. Generally, different members of the team should take responsibility for interviewing different individuals in different locations. However, while rational in terms of time and travel, such divisions of labour require adherence to the discipline of writing-up and sharing interview records among Team Members. The use of laptop computers, e-mail and free-form databases is potentially a very effective means of sharing such information and enabling all team members to contribute to and benefit from the process of constructing the narrative and 'baseline'.

As noted above the strength of multi-disciplinary teams lies in the differing perspectives that can be brought to bear on the issues and it is vital that the team has sufficient opportunities to get together at different stages of the evaluation process to discuss overlapping issues and conclusions.

Emergency personnel often have an extremely high degree of personal involvement and engagement in their work. This can make them very defensive when challenged or questioned about their achievements and performance. Evaluators need to be aware of this and be sensitive in their approach. Involving regular programme staff in evaluations of other programmes can help considerably in this regard. Programme staff not only understand what others have been through, but are often well-placed to ask the right sort of questions. At the same time, the next time they are in the position of being evaluated, they can better understand the process and use the experience constructively.

Authors should be able to stand by their work. In one recent evaluation, the chapter authors never saw the final draft before it went out. When objections were raised to some of the findings, the authors denied being responsible for them! An evaluation is part of a dialogue, and this cannot be effective if those involved cannot claim the work as their own.

A list of all the people interviewed should be provided as an annex.

Innovation and methodological pluralism

'When so much is unknowable and so unpredictable, solutions have to be sought through methodological pluralism, through flexible and continuous learning and adaptation, and through the exercise of judgement' (Chambers, 1992).

Evaluation in emergency environments is as much an art form as a science. Evaluators need to be flexible, to adapt standard tools, to be innovative and to use a mix of tools. Good use needs to be made of any appropriate secondary data sources.

The need to be opportunistic

Evaluators need to be opportunistic, to collect information, in whatever form, wherever possible. They need to collect information from several different perspectives, in order to check the validity of information, improve accuracy as a counter to the lack of extensive sampling.

10

Sources of information and information-gathering techniques

10.1 Preliminary data collection

Preliminary data collection, before field-work begins, is a vital part of evaluation. A team member, with good awareness of the main issues, needs to begin to work out what are the key data sources: for example, reports and correspondence produced by the main agencies, and then to find out where this information is held and how it is organised. The job of sifting through files can be extremely time-consuming, however. It is generally much more efficient to start with basic interviews with a small number of key personnel who can direct the evaluator to the right material. It is important, however, to ensure that biases are not introduced at this stage that may increase attention given to certain issues, at the cost of attention given to others of equal importance (Wood, 1998).

It is also important not to forget about data available from international institutions, such as WHO, UNDP and the World Bank, to name just a few. There may also be academic works available on the area where the evaluation is to take place.

10.2 Interviews

In the evaluation of humanitarian assistance programmes, as discussed above, interviews are generally the most important source of information on what happened and why. Evaluators need to talk to as wide a range of ‘stakeholders’ as possible, to build up as complete and balanced picture of an intervention as possible: agency staff, local NGO partners, host government officials – national and provincial – UN officials. Local staff will generally have differing perspectives to HQ staff, and communication between the two can be an important factor in shaping a response. The assessment of the lines of communication between the two sets of staff, can reveal where there is an absence of shared analytical tools (see Box 10, opposite). It is also helpful to talk to individuals who do not have a direct role or stake in the implementation process, such as local academics or journalists, as NGOs and UN and Government officials may have hidden agendas, not always immediately apparent to an (external) evaluator.

Evaluation interviews should, ideally, be relaxed and undertaken in a spirit of a shared desire to learn from past experience. Obviously, to some degree, this is easier said than done. However, the evaluator needs to take into account that the person being interviewed has probably done their best, often in difficult circumstances. This does not mean that the evaluation is not a critical appraisal, merely that it is not inquisitorial. Where mistakes have been made, the evaluator needs to try and assess the context in which they were made: what were the decision-making processes; and, what information was available at the time the decision was taken.

Box 9

Joint consultations with stakeholders

In an evaluation of EU-funded aid programmes in Ethiopia, the evaluators convened a workshop for programme stakeholders – NGOs, donors, UN agencies, EU advisers and Government officials – which developed a matrix comparing the EC with other donors (Maxwell, 1997). A set of desirable attributes of a donor was generated, and the participants ranked the Commission in relation to other donors. This was found to be a helpful exercise in discovering the key issues that confronted those in country in their attempts to work alongside the EC.

Box 10

Talking to all the stakeholders.

An evaluation of an emergency intervention in Turkana (Jaspars et al, 1997) found that the food security indicators – developed over a number of years with local inhabitants – were not able to reveal the onset of acute food insecurity, despite local staff being aware of increasing problems. Some of the indicators moved in a direction normally consistent with an improvement in food security, while others were unable to distinguish between coping strategies carried out all the time, and those used only during times of stress. Data showing a more or less steady increase in malnutrition was difficult to use without an understanding of the normal seasonal pattern of malnutrition in the area. As a result, despite a worsening situation, local staff could not convince HQ staff of the need for an emergency intervention on the basis of the food security and nutritional information available. Indeed, the data available did not show that food insecurity was acute. This was partly due to the fact that “Food security indicators (had) not changed since 1988, but at least part of the population (had) undergone a fundamental change in livelihoods over that period”.

10.3 Working in areas of conflict and instability

Evaluators need to be aware of how difficult it can be for those caught up in conflict and instability to discuss what they have experienced. Regardless of who the team is interviewing (be it officials within the government or a ‘rebel’ faction, relief workers, military personnel within the factions and in peacekeeping contingents, or civilians within the affected population who received assistance provided), it is important that they bear in mind, and as far as possible empathise with, the experiences that their interviewees have endured. Affected populations who have just experienced a civil war may have been forcibly displaced, had relatives and friends killed, either in the fighting or during atrocities committed against civilians, and seen their personal, social and perhaps cultural identities shattered. Psycho-social trauma may affect much larger numbers than is evident to an outsider, particularly if unfamiliar with the local language and untrained in the diagnosis of post-traumatic stress disorders. Exposed to such extreme

experiences and perhaps having themselves witnessed massacre sites, it is not unheard of for members of evaluation teams to also experience mild forms of traumatic stress disorders¹⁴.

Civil wars greatly exacerbate and deepen existing fault lines within a society and may well create new cleavages in previously coherent groups (e.g. between those who stayed through the conflict and those who sought safety outside the country; those who came to support a new faction and those who remained loyal to a former government or faction). Such is the intensity of feeling and the polarisation of affected societies that objectivity and the truth become difficult concepts to uphold. It is not uncommon for evaluators working in complex emergencies to experience two intelligent and articulate adults giving completely contradictory versions of the same event (see Box 11, below). In such situations, evaluators may have to accept that their search for ‘the truth’ may never be successful. The implication of this is that the veracity of information collected cannot be taken for granted and will require constant checking and cross-checking. When writing the report care will need to be taken to ensure the careful wording of particularly sensitive sections.

Box 11

Examples of different, and intensely-held, perspectives

The Joint Evaluation of Emergency Assistance to Rwanda was presented to a three day, Government sponsored Conference in Kigali in September 1996. The finding in Study 3 that international assistance had been biased towards the refugee population and that 60% of all assistance provided to the region had gone outside Rwanda was hotly disputed by Government officials who claimed that their own analysis showed that the bias was much worse with 95% of the assistance having gone outside the country. The evaluators were themselves accused of being biased towards the refugees and against the new Government.

10.4 Gaining the beneficiaries perspective

Humanitarian assistance is essentially a ‘top down’ process. Of necessity, it often involves making assumptions about assistance needs and the provision of standardised packages of assistance. Even where time and the situation

permits, humanitarian agencies are often poor at consulting or involving members of the affected population and beneficiaries or their assistance. Consequently, there can often be considerable discrepancy between the agency's perception of its performance and the perceptions of the affected population and beneficiaries. Experience shows that interviews with beneficiaries can be one of the richest sources of information in evaluations of humanitarian assistance. Interviews with a sample of the affected population should be a mandatory part of any humanitarian assistance evaluation.

Consulting beneficiaries is not only a practical and effective way of assessing impact, it also gives a voice to those who may have lost their usual communication channels. Such 'downward' accountability is an important component of evaluation, particularly where state structures are weak or non-existent, and where access to the press or local antagonism to journalists can mean that much of what goes on is unrecorded and unobserved by outside forces. In such situations, aid agencies can yield enormous power, and effectively be accountable to no one. Evaluations can go some way towards redressing such a state of affairs. Personal experience has shown that beneficiaries generally relish the chance to be given the opportunity to express their feelings, giving the participatory process (provided comments are taken seriously and followed through by programme management) benefits above and beyond that arising from the actual information gathered.

The use of Rapid Rural Appraisal (RRA) and Participatory Rural Appraisal (PRA) techniques offer an important way in which to understand beneficiary perceptions. They can help to 'fine-tune' quantitative information that has been aggregated over relatively large areas or population groups and increase the focus on those groups who have been most severely affected (Mitchell and Slim, 1997). There is now a considerable body of literature on the various tools of PRA and RRA, though little specifically targeted to the humanitarian worker. The attitude of the researcher is one of the main determinants of successful use of such techniques: the willingness of the evaluator to be flexible, adaptable and prepared to learn from those whom the project is there to assist is essential.

PRA and RRA techniques have become popular partly as a result of increasing frustration with the use of more 'developed' statistical techniques in development work. Large-scale surveys were often found to be expensive, often of poor

Box 12

The gap between theory and practice

Kruse et al (1997), in their NGO evaluation synthesis study found that ‘while involvement with beneficiaries frequently occurred, it was often in an artificial, perhaps strained context where hurried answers were sought to hurried questions’. The same study also reports on the ‘wide gap between theory and practice: while almost all NGOs speak of the importance of participation, there is a paucity of evidence of participation in NGO evaluations.’ A recent study which aimed to seek out and analyse examples of successful participatory practices concluded that examples were extremely rare (Martin and Quinney, 1994). The NGO Synthesis Study, in its annex on Kenya finds that ‘though there is currently plenty of talk about participation and participatory methods, there is a paucity of evidence of participatory processes on the ground’.

design, and producing information of questionable significance and relevance. Furthermore, the time taken to train enumerators, carry-out surveys and process the results can take years, and be of little use to those needing to make more urgent policy or practical decisions. In emergency situations, it is likely to be even more difficult, if not impossible, to carry out large surveys capable of generating statistically significant data. However, the difficulties of carrying-out genuinely participatory work in such situations should not be underestimated¹⁵:

- in complex emergencies, where political sensitivities can be acute, beneficiaries could be exposing themselves to risk by talking to evaluators. In addition, by talking to one group and not another, agencies can be perceived to have “taken sides”, thereby compromising their neutrality and independence;
- following conflict, traditional communities may have been broken-up. This can make it difficult to find genuinely representative interlocutors. Mitchell (1997) talks of the problems of trying to identify an ‘entrance point’ into a community in Mozambique. After decades of war and massive population displacement, the notion of legitimate authority had been eroded. People were not only distrustful of outsiders but were living atomistic lives, without

recourse to a set of commonly held beliefs and reciprocal ties. The lack of an ‘entrance point’ – such as a village chief or group of elders made it very difficult to begin a participatory planning process;

- it may be unacceptable, given the traumas that some populations have gone through, to ask them to relive their experiences for the sake of an evaluation;
- it is important not to raise expectations that cannot be met. Communities, or groups of beneficiaries may not understand the evaluation process, and cooperate in the belief that more aid is being offered when, it could transpire that the evaluation results in a cutting of aid.

A combination of interviews with individual households, women’s groups and open group discussions involving men as well have proven to be very productive in some contexts. Given the problems outlined, however, ensuring the confidentiality of interviewees can sometimes be necessary. Evaluators need to ensure that they do not speak only to local leaders (who may have their own agendas, and not represent more vulnerable groups), but that a broad cross-section of the population is included – the impact of programmes may have very different effects on different categories of people. A reading of the PRA/RRA literature can reveal many of the biases that can slip into such work

Box 13

Talking to beneficiaries

During the height of the war in Mozambique, a participatory exercise was carried out over a few days. While beneficiaries were pleased to have received food aid and seeds, they were unhappy with the fact that they had only received one hoe per family. Even during periods of violent conflict, families depended a great deal on being able to carry out work in the fields for those who were better off. With only one hoe, this was impossible, as they also needed to cultivate their own fields. The participatory exercise helped agencies understand the extent of coping mechanisms and trading arrangements, which many had assumed had collapsed. At the same time, the school rehabilitation, strongly promoted by the government, was not supported by the local population. They were far more interested in more school books and teachers than in new buildings.

where there is not an awareness of how communities organise themselves. Wherever the target population, whether in camps or dispersed, there will be power structures at play that, unless understood, can lead to a false picture of the situation being gained. It is important that material obtained through PRA/RRA methods (or any other method, for that method) is cross-checked, using other sources of information, or through direct observation.

The deliberate seeking out of those who did not benefit from the assistance available can also be fruitful as it may reveal problems with the targeting and beneficiary selection processes used by the agencies. In addition, it is important to speak to representatives of the host population, to identify if their lives have been affected by the arrival of displaced or refugees. Ideally, anthropologists, or others familiar with the culture and the indigenous language would undertake such work.

On some occasions there will not be the time available to carry out a comprehensive PRA exercise, with high levels of participation. However, there should always be an attempt to use some PRA/RRA methods: semi-structured interviews with key informants, focus group discussions, and limited direct observation should be considered minimum requirements of any evaluation. However, it is important to be aware that RRA was originally conceived as a 'rough and ready' information-gathering approach, and not the semi-discrete research method in its own right to which some have elevated it. Any results from PRA/RRA exercises need to be used with the same caution that would be exercised with information from any other approach.

Box 14

Understanding local power dynamics

During the Joint Evaluation of Emergency Assistance to Rwanda, the two anthropologists on the team found that the important members of the former regime controlled all the strategic areas in the camp – around the principal entry points, and along the main access routes. All the camp inhabitants were monitored by those close to members of the former regime, and were not free to speak openly about what was going on in the camps.

Analysis techniques

11.1 Impact

A review of aid to Afghanistan (Donini, 1996) concluded that “the problem with discussing the impact of international assistance on the people of Afghanistan is that none of us can really claim to know what the impact has been. It is tempting to believe that with time the true impact will become clearer but this may be just wishful thinking.....Some of the changes in Afghan society may be the result of conflict, some the consequence of international assistance, while others may simply be natural developments which would have come about anyway. It is difficult, if not impossible, to distinguish between the various causes and effects. Aid programmes may have accelerated some of these processes or, equally, they may have retarded them”.

The difficulty of measuring impact is certainly not a characteristic unique to humanitarian aid programmes. Indeed, much of what is said in this section could apply equally to evaluations of development aid. Whether for humanitarian or development aid, the three principal challenges for those wishing to measure the impact of an humanitarian intervention are:

- i) a lack of consensus on what constitutes a desirable outcome;
- ii) the problems of data collection and data availability;
- iii) attributing impact to an intervention.

A lack of consensus on what constitutes a desirable outcome

While technical performance standards are being developed for the humanitarian world (see section 11.3 on page 90), there remains a lack of consensus about the rationale for and scope of humanitarian assistance: saving lives is a primary aim of most emergency programmes, but few would accept that a programme can be measured purely on the basis of changes in mortality or morbidity rates. Quality of life issues are also important. Protection from arbitrary detention, intimidation and physical violence is increasingly becoming an objective of humanitarian assistance programmes.

Where a range of outcome indicators are selected, how should these be ranked? How does one deal with the potential trade-off between long-term and short-term impacts? Can the provision of clean water be justified as a cost-effective approach to lowering morbidity in the longer-term, or is access to clean water a right in itself? Should just enough food aid be given to ensure minimal physical survival, or should it be in sufficient quantities to make sure that beneficiaries are able to live in a healthy manner, without destroying fragile local environments in the search for food, or running down household assets and jeopardising longer-term survival? What happens when food aid damages incentives for local farmers?

The problems of data collection

Collecting data, whether on output or outcome indicators, can be extremely difficult in an emergency situation, although some activities are far more amenable to measurement than others. The tonnage of food delivered to a refugee camp, for example, is easier to monitor and record than a more intangible benefit such as the provision of health services. However, even for the same activities, the availability of data can vary considerably depending on the area of operation, the agency involved and the particular beneficiary groups.

Differences in data availability also depend upon the number of organisations involved (see Box 15, opposite). Most of the food aid supplied to the refugee camps was delivered by a single UN agency (WFP), whereas there were up to a dozen NGOs working in health care in some of the camps. It was thus far easier to track down sources of data on tonnes of food delivered than on vaccinations given or consultations carried out. The number of organisations involved also made it difficult, even where data was available, to isolate the contribution of any one particular organisation.

Box 15

Differences in data availability

The Study III Team of the Joint Evaluation of Emergency Assistance to Rwanda found that information on refugee operations outside Rwanda was more plentiful and easier to obtain than information on those operations inside Rwanda that provided assistance to IDPs and other beneficiary groups. This experience largely reflected the fact that a single agency, UNHCR, had an unambiguous mandate to protect refugees and coordinate activities in the camps outside Rwanda, whilst inside Rwanda, coordination and the division of responsibilities was much more confused. Initially, this confusion was a result of the civil war and the high levels of insecurity in much of the country. Later it was attributable to the combination of a new, inexperienced and poorly-resourced government in Kigali, and the lack of a single UN organisation with a mandated responsibility for IDPs and strong enough to ensure standardization in data collection techniques. Amongst the refugee sites, Goma, with its larger scale and higher profile, had the best epidemiological information available, thanks to the presence of teams from CDC, Epicentre, WHO, and UNICEF.

In the early stages of an emergency operation, the heavy demands upon relief personnel and organisational structures can preclude adequate data collection. UNHCR, for example, only had two international staff present in Goma on the first day of an influx which, over the space of a few days, saw the arrival of approximately 850,000 refugees. Outside a camp setting, it can be difficult, if not impossible, to establish population size, and mortality rates simply may not be calculable.

Attributing impact¹⁶

Impact analysis involves not merely examining whether the project's objectives have been achieved but also the degree to which the outcomes achieved have arisen as a result of the project. This requires the evaluator to separate the contribution of an aid intervention from the contribution due to other factors, in other words, to estimate what would have happened in the absence of the intervention (the counterfactual). In an emergency setting, even in the relatively

controlled environment of refugee camp where data on mortality and morbidity can be collected, such causal analysis can be extremely difficult. For example, following a cholera epidemic, mortality rates will fall regardless of whether there is an intervention or not. Similarly, improvements in nutritional status of a target population may be due to the re-opening of a road to an important market, rather than a food aid programme.

There are three common approaches to dealing with the problem of attributing impact:

- i) the scientific approach
- ii) the deductive/inductive approach
- iii) the participatory approach

The scientific approach

The scientific approach is more favoured by those wishing to generate quantitative measure of impact, for which there is a strong demand, as such information is more easily analysed, packaged and presented than qualitative information, and hence of more use for those who need to show to donors and parliaments that interventions are having an impact. Furthermore, quantitative measures of impact (given information on the costs of interventions) are needed to calculate the cost-per-‘unit of impact’ (the cost-effectiveness) of an intervention – important for managers having to decide where to allocate scarce resources to ensure maximum impact.

In most of the social sciences it is impossible to carry out pure ‘scientific’ experiments to determine impact, and so quasi-experimental techniques are used. One such method for this is multiple regression. The enormous data demands (as well as other disadvantages) of such an approach, however, rule out its use (except in the most exceptional circumstances) for humanitarian assistance programmes. Another method, commonly used in many research fields, involves the creation of control groups, so that changes over time between those affected by a project and those outside a project can be compared. In an emergency programme, however, the use of control groups is neither practical nor morally acceptable: one cannot deliberately exclude one group from a relief programme solely in order to measure the impact of the programme. While it does sometimes happen that certain groups are excluded for various reasons –

lack of access, lack of agency capacity or lack of knowledge as to the location of all vulnerable groups – comparisons between those who received assistance and those who did not need to be used very carefully. Mortality in the two groups may be equal, but the group not receiving aid may have had to sell important assets, thereby severely compromising longer-term survival strategies. Without a deep understanding of what is going on, accurate conclusions will be very difficult.

As a less rigorous alternative, before and after comparisons can be made of the same group. Ideally, this requires the collection of base-line data, so that conditions before the intervention can be compared with conditions afterwards. Again, in an emergency setting, the data is not always available, and there is not always time to collect it before beginning an intervention. Where communities are undergoing great changes, such as during a war or where populations uproot and flee to another area, there will generally be too much uncertainty to attribute impact to an intervention whether before-and-after data exists or not.

The deductive/inductive approach

The deductive/inductive approach, on the other hand, involves a more anthropological and socio-economic approach to determining impact. The approach relies heavily on interviews with key informants, visual observation and inductions of lessons from other cases that are similar or comparable. It does not require a statistically valid ‘proof’ of impact, but seeks to provide a narrative interpretation of the links between an intervention and any impacts that have a high degree of plausibility. It understands that there are often several ways of interpreting impact, and that interpretation requires a balanced approach. The validity of a specific report has to be judged by the reader on the basis of the logic and consistency of the arguments; the strength and quality of the evidence provided; the degree of cross-checking of findings (triangulation); the quality of the methodology and the reputation of those involved. It should draw upon academic and operational research, and the development of conceptual frameworks that enhance understanding of how aid interventions work.

With this approach causal links between interventions and final impact are deduced by a combination of information from key informants and from the evaluators own sense of how the world works.

The Participatory approach

Participatory approaches to impact measurement depend, to a greater or lesser extent, upon obtaining the views of those most directly affected by a programme. The most radical proponents of such approaches are those linked to what is known as Fourth Generation Evaluation. Proponents of this school completely reject the assumption that the ‘true’ nature of external reality is discoverable through the scientific method. Instead, they argue that the role of the evaluator should be one of facilitating dialogue among a wide variety of a programme’s stakeholders, the objective being to attain consensus upon the programme’s value and contents (Fishman, 1992). The scientific method is rejected as empowering the elites, and ignoring the complexity and diversity involved in survival strategies. Such radical approaches do not resolve the problem of attribution, rather they sidestep the problem by questioning the right of outsiders (managers, evaluators) to set programme objectives in the first place. In this approach the evaluation exercise itself is empowering: it does not merely seek to measure the attainment of objectives, but is one of the processes by which such objectives are attained.

In practice, most evaluators use a mixture of approaches, with the deductive/ inductive approach being the most common. Genuinely participatory

Box 16

Disaggregating data

It is vital that aggregate data is not used in isolation as a way of measuring performance. Experience in the camps in Goma in 1994, for example, revealed that aggregate supplies of food reaching the camps were adequate. Indeed, the Joint Evaluation found that the WFP logistics operation was generally very efficient. However, when food-basket monitoring was carried out, it was found that certain groups within the population, such as households headed by a single mother, were not receiving the amounts they were due. Local power structures, not understood by aid workers, were influential in determining how aid was distributed, and resulted in high levels of malnutrition amongst certain vulnerable groups in the camps in Goma. It was only when distributions were targeted at individual households rather than at larger groupings, that nutrition indicators improved amongst the vulnerable.

approaches, involving beneficiaries, are complex and time-consuming, and subject to a range of difficulties. While it is essential for evaluators to talk to beneficiaries, there are often significant limitations to what they can say about overall programme strategies. Even where impact can be shown, the programme may still not be the most appropriate. For example, data on weight growth could show that a supplementary feeding programme is running well, but this may be at the cost of the health and nutrition status of other family members as a result of the time the child's carer is spending attending the programme. Alternatively, a water programme could have a greater impact by preventing overcrowding and disease (a major determinant of malnutrition) and, by increasing supply, freeing up family time to forage for food and care for infants.

For accountability purposes, and for reasons of clarity, evaluators should make sure that the link between their findings and the evidence used is clear. Output data should be presented, whenever available, not only for accountability purposes, but also to allow the reader to make his or her own judgement as to the overall findings of the evaluator. The logic of the evaluator should be clear and convincing. At the same time, it must not be assumed that evaluators are infallible – methodological clarity enhances the accountability of the evaluator to those being evaluated.

11.2 Cost-effectiveness

Evaluations should comment on the impact of humanitarian aid programmes, and not focus solely on monitoring the efficiency of project implementation. However, humanitarian programmes often take place in complex, confused circumstances, where plans change regularly, where information is scarce and where little may be predictable. As a result, qualitative and deductive methods of measuring impact, that involve beneficiaries, are likely, in general, to be more appropriate than methods that seek to be 'scientifically rigorous'.

Cost-effectiveness analysis seeks to determine whether the objectives of a programme could have been achieved at lower cost (or whether greater impact could have been achieved for the same cost). It is not the same as cost-benefit analysis, as it does not require the valuation of benefits, making it more appropriate for programmes with social development or humanitarian goals. Cost-effectiveness is a broader concept than efficiency, in that it looks beyond how inputs were converted into outputs, to whether different outputs could

have been produced that would have had a greater impact in achieving the project purpose. In other words, cost-effectiveness analysis seeks to determine the least expensive way of realising a certain level of benefits, or a way of maximising benefits per unit cost.

Cost-effectiveness analysis is regularly required of evaluators, though, in practice, there are usually severe limitations on the scope of such analysis. Cost-effectiveness analysis requires not only the quantification of outcomes, but also the quantification of costs, and analysis of the link between the two. All the difficulties associated with the quantification of impact – lack of consensus as to what constitutes a desirable outcome, difficulties of data collection and attributing impact (discussed above) are, therefore, also associated with analysis of cost-effectiveness. In addition to these not inconsiderable problems, are a range of other difficulties, both practical and methodological.

The difficulties of obtaining cost data¹⁶

There are numerous difficulties involved in obtaining data on costs:

- i) many agencies do not record their expenditure by activity or beneficiary group, particularly those working on multi-sectoral relief programmes. Their accounts are primarily produced for auditing requirements, and do not to allow for detailed cost-effectiveness analysis of projects;
- ii) where expenditure is classified by activity, no standard classifications are used, so the same project can be recorded in different ways by the NGO, UN agency and donor involved. In addition, the same project can be recorded using different currencies and under different financial years, as donors fund NGOs from countries different to their own. This makes it very difficult to reconcile information received from multiple sources;
- iii) the value of aid-in-kind is measured in a number of different ways;
- iv) complex sub-contracting arrangements are sometimes entered into between and amongst bilateral agencies, UN organisations and northern and southern NGO partners, involving the transfer of human and financial resources, as well as aid-in-kind. This makes it difficult to trace the flow of funds, let alone work out end-use of resources;
- v) there is no standard way of dealing with the issue of overheads.

Box 17

Counting all the costs (and all the benefits)

It is important to consider not only the costs borne by the donors, but also by beneficiary and host populations. For example, the supply of slow-cooking beans and whole grain maize instead of flour has led to intense deforestation around refugee camps, this cost borne almost solely by host communities. Similarly, centralised health provision, for example, may appear to be the cheapest option for a donor, but may involve beneficiaries walking considerable distances. Costs for all involved need to be included in any analysis.

It is also important to count all the benefits. It may not be appropriate to merely count the latrines built in a sanitation programme: skills and knowledge, and community structures may also have been left behind and should be considered.

The context-specific nature of most cost-effectiveness information

Although many relief operations involve undertaking essentially the same activities, such as the provision of food aid, the setting up of airlift capacity and the emergency provision of water and health services, each relief operation is unique to the circumstances in which it takes place. Cost-effectiveness analysis of one particular relief activity or operation will, therefore, tend to produce information that is highly context-specific.

Experience of evaluating one of the key interventions in Goma, namely the provision of water to Kibumba camp, demonstrated this point. At the early stages of the relief operation, water was pumped out of Lake Kivu by a private contractor into water tankers operated by a range of NGOs, UN agencies and donor teams, which then delivered the water to tanks/reservoirs built and run by OXFAM-UK, which also constructed and operated the reticulation systems within the camp. The cost of producing a litre of water in this context bore almost no relation to the cost in any other refugee setting or, indeed, even in neighbouring camps in Goma which had quickly moved to systems drawing on nearby surface water sources.

Even the more common example of water provision by borehole and hand-pump cannot be adequately described without detailed information on the

particular circumstances involved: the depth to be drilled for boreholes; the nature of the layers of rock between the surface and the aquifer; the quality of the water being produced; the distance recipients needed to walk to be able to collect the water, the time needed to queue, etc.

It is also important to be aware that, as one seeks to reach more of the population, particularly those in more isolated areas, the cost per beneficiary will generally rise. Cost-effectiveness analysis does, therefore, need to be used with care. It should not, for example, be used to compare a programme in one country with a programme in another country. It should also be remembered that a cost-effective programme is not necessarily a cheap one.

Cost-effectiveness or cost-efficiency?

In a world where numerous and competing demands are being made upon finite aid resources, it is essential that the relationship between costs and performance be fully examined. Too many evaluations talk about objectives and make recommendations in the absence of any sense of the costs associated with alternative approaches. Despite limited experience of successful cost-effectiveness analysis (see Box 18, opposite), the difficulties described above mean that the more limited concept of cost-efficiency analysis is likely to be more appropriate. ‘Efficiency’ refers to the process of implementation of a project, in contrast to ‘effectiveness’, which refers to outcomes. A cost-efficient project is then one where the inputs are supplied and used in such a way as to achieve certain specific outputs at lowest cost¹⁷. Given that many humanitarian programmes involve the provision of large quantities of material inputs – food¹⁸, seeds and tools, domestic items, shelter material – such analysis has the potential to highlight huge savings. (For example, see Section 3.2 on page 30) which discusses the findings of cost-effectiveness analysis carried out during the Rwanda evaluation (Borton et al, 1996.)

Box 18

Examples of cost-effectiveness analysis

MSF carried out a study of the cost-effectiveness of an emergency programme to combat Visceral Leishmaniasis in southern Sudan, which showed that, despite the high cost of transport, personnel and treatment, in terms of years of life saved, the intervention was extremely cost-effective (Griekspoor, 1998 forthcoming). MSF also undertook cost-effectiveness analysis of an operation to combat human trypanosomiasis in northern Uganda (Perea and Paquet, 1996). The evaluation recorded that 5,000 people were treated who would otherwise have died, and that this was directly attributed to the MSF programme, as no other control activities were implemented by other parties before or during the intervention. MSF were able to calculate the cost per case treated and the cost per death averted.

Both these examples relate to vertical control programmes to combat diseases with particularly high mortality rates, where reductions in mortality were attributable solely to the intervention. However, most diseases are not as fatal, nor as amenable to treatment. Reducing diarrhoeal deaths – often the principal cause of death in emergencies – requires a range of inputs, including water and sanitation, oral rehydration, medical treatment, health education and a range of nutritional support programmes. No single dose of a drug will ‘cure’ diarrhoea, making it much harder to isolate the impact achieved by a humanitarian agency from a range of other factors, such as improved rains, improved access to markets, or any other event that could lead to an improvement in health status.

11.3 Standards, norms and benchmarks

Against what does one measure performance? What is an acceptable level of ‘coverage’? How do you measure ‘coherence’? If we say a project has a positive impact, what do we mean: in terms of a recognised standard, in terms of the subjective opinion of the evaluators, in terms of what the project set out to do, or in terms of the opinions of the target population? What happens when use of these different bases of measurement produce contradictory results? What if refugees are pleased at the volumes of food aid received, but the evaluators consider that the amounts were too large and damaged local markets, or the donors believe that the amounts were more than were needed and hence the programme cost more than necessary?

Such issues have long been a cause of concern for evaluators: the absence of widely accepted benchmarks or a clear legal framework for humanitarian assistance has increased the difficulties of evaluating humanitarian assistance. In the past few years, however, there has been a considerable effort by the humanitarian community to strengthen the framework in which humanitarian assistance programmes take place. This effort has led to the development of a widely-accepted Code of Conduct (see Box 19, opposite) as well as the development of technical standards through the SPHERE Project (see Box 20 on page 92), which should be of considerable value to evaluators.

These basic standards, once they have been developed, will provide a benchmark against which performance in certain sectors can be measured. However, while of use for the more technical relief activities, they do not reduce many of the challenges involved in the evaluation of humanitarian assistance programmes. How does one evaluate performance where, due to problems of access, political will or funding, it was not possible to reach the required standards? Equally important, the SPHERE standards will be of less use for the less ‘delivery-focused’ objectives of humanitarian assistance projects, most notably the protection of citizens in conflict areas. Though not often stated explicitly, in many situations this may be the key objective of a humanitarian assistance programme.

It is important, however, that the development and use of standards does not take attention away from the beneficiary perspective. Involving the beneficiaries should remain a fundamental component of the evaluation process.

International Human Rights and International Humanitarian Law appears to offer some sort of framework for the conduct of those involved in providing assistance. However, it is important to be aware that “the rights regime, taken as a whole.... is designed more to protect against threats to security (physical, economic, social, political) than to meet people’s immediate needs where they cannot do so themselves” (Darcy, 1997). Furthermore, while the need to remove the factors constraining people’s freedom of action may be the paramount humanitarian concern in any given situation, “it is important to locate responsibility for protection where it belongs During armed conflict, it is

Box 19

Code of conduct for the International Red Cross and Red Crescent Movement and NGOs in disaster relief

1. The humanitarian imperative comes first
2. Aid is given regardless of the race, creed or nationality of the recipients and without adverse distinction of any kind. Aid priorities are calculated on the basis of need alone
3. Aid will not be used to further a particular political or religious standpoint
4. We shall endeavour not to act as instruments of government foreign policy
5. We shall respect culture and custom
6. We shall attempt to build disaster response on local capacities
7. Ways shall be found to involve programme beneficiaries in the management of relief aid
8. Relief aid must strive to reduce future vulnerabilities to disaster as well as meeting basic needs
9. We hold ourselves accountable to both those we seek to assist and those from whom we accept resources
10. In our information, publicity and advertising activities, we shall recognise disaster victims as dignified humans, not hopeless objects

... likely that beyond people's own efforts to help and shield each other, only the warring parties can provide effective protection" (Darcy, 1997).

Evaluators should be conversant with recent writings on humanitarian policy and practice, and, in the evaluation report, reference should be made to key works. This can help locate the evaluation's findings in a widely discussed framework, and can give the evaluation greater credibility than would be achieved if the report is seen simply as a subjective view of one individual¹⁹.

Box 20

The SPHERE project

Over the past year, the SCHR, a coalition of European NGOs involved in humanitarian action, in collaboration with InterAction, a US coalition of NGOs, has been developing basic standards of practice for the supply of food and water, the provision of shelter and the delivery of medical services to disaster victims. These standards look not only at what is delivered, but at how it is delivered, and how the providing agency interacts with the beneficiaries and its donors. Issues of needs assessment, targeting, accountability and evaluation are examined along with cross-cutting issues such as gender and environmental considerations. The standards have been developed by considering the beneficiaries' needs and rights, rather than taking a supply-driven approach. The question asked has been "what do those who suffer in disasters have a legitimate right to expect from the humanitarian system in terms of the quality of service provided?"

Web address: <http://www.ifrc.org/pubs/sphere>

12

Writing the report, preparing recommendations and presenting findings

12.1 Writing the report

The writing of the final report, its structuring and language, is pivotal to the whole credibility, competence and independence of the evaluation. It is important that Evaluation Managers appreciate the degree to which the process of writing and analysis are entwined. Adequate time must be allowed for the writing process. A useful rule of thumb is to allow 50% of the time spent in the field by each member of the team.

A draft report should always be prepared and circulated for comment to those organisations and individuals involved in the evaluation (where possible, including all those interviewed). Given the complexity of the subject, the inadequacy of documentation, the fallibility of evaluators and the issue of whether an objective truth exists in the context of civil wars, it is vital for the credibility and competence of the process that a draft is shared widely, that adequate time is allowed for it to be considered and for comments to be received. Experience shows that four weeks is an appropriate time to allow for comments on a draft report. Whilst some comments may be easily dealt with others may require additional investigation. Clearly, not all comments can be incorporated into the final draft. Where these concern major criticisms of the report, however, the evaluators should make clear, in the report, why the criticism is rejected.

(Where the report is to be placed in the public domain, agencies that feel unfairly criticised in the evaluations could be allowed to add comments in an annex.) Adequate time should be made available for the evaluation team to review all comments. It is common to find that the sharing of the draft actually ‘smokes out’ additional information that had not been shared earlier in the process due to the agencies’ perceptions of the sensitivity of the information.

It can be useful for an outline of the report to be drafted at an early stage in the process, after the initial data collection and interviewing has taken place, but before the field work. This can allow evaluation teams and management to ensure that both are happy with the direction that the evaluation is taking.

It has already been noted that one of the strengths of a multi-disciplinary team is the differing perspectives it can bring to bear on issues. All team members should, therefore, be involved in discussing the findings and linking these to conclusions. Tensions which may arise between the team leader and individual subject specialists on the nature of the conclusions can be managed more effectively if the whole team is brought together to discuss these. Ideally, the team should hold workshops to discuss preliminary findings on return from fieldwork (where they have not been working together in the field) and then subsequently to discuss comments received on the draft report and any new information provided by the agencies.

Whatever its scope or nature, an evaluation report will maximize its potential impact if it presents its findings, conclusions, recommendations and follow-up sections separately. If readers disagree with the recommendations (or find themselves unable to implement them because of political constraints), they may be able to agree with the findings or conclusions.

Comprehensive discussions within a steering committee, or management group, on the draft report may contribute to a sense of ownership of the report on the part of the management group.

In any large-scale evaluation, conflict over the nature of the recommendations is probably inevitable. In preparing recommendations, in order to minimise such conflict, there needs to be a clear link between the recommendations and the evidence in the body of the report to support them.

The form in which recommendations should be made is the subject of continuing debate among evaluation specialists. Some would argue that evaluation reports

should contain specific, implementable recommendations detailing the actions agencies should take in order to improve future performance. Such recommendations might also spell out who is responsible for implementing each recommendation and who is responsible for monitoring whether this action takes place. Others would urge caution, favouring findings and conclusions over specific recommendations, so as not to burden policy-makers with recommendations that could lead to unimplementable policies. If recommendations are required, an evaluation team might provide policy-makers with options rather than a single recommendation, together with an analysis of expected consequences. Different issues may require different responses. Technical issues may lend themselves to specific recommendations in the final report. In dealing with broader issues, however, it may be more productive to deliver the analysis to a workshop of decision-makers and evaluators which negotiates follow-up action.

Regardless of whether recommendations are negotiated or independent, specific or general, if evaluation reports are to do more than gather dust on the shelves, they must be timely. In spite of the complexity of the issues involved, the speed at which consciousness of the need for emergency assistance fades means that evaluations must take place promptly. Organisations are easily able to ignore evaluation reports which are published two or three years after the events to which they refer.

At the end of the field visit of evaluators, there should be a workshop where the main findings are presented to the stakeholders. The WFP Liberia evaluation found that “mission debriefings contributed to mutual understanding enormously” (Apthorpe, 1996).

12.2 Structure of the evaluation report

Evaluation reports should include a brief executive summary, ideally published separately from the main report, but informing readers where a full copy can be obtained. Policy makers and senior managers rarely have the time to read a full report, and there is a need to cater for this important audience. This should be followed by a short section explaining the methodological approach used, followed by a description of the context of the assistance.

Evaluations are of potential interest to agency managers, beneficiary representatives, local governments, donor officials, students of humanitarian

issues, the press, as well as the general public. Ideally, each audience should receive a different products tailored to their needs. While the availability of resources for publication will generally limit what can be done in this respect, it is important to underline that attention to the dissemination of evaluation results is as important as carrying-out the evaluation in the first place.

Well-executed and well-presented evaluations are a valuable resource for the humanitarian system as a whole, even where the evaluation concerns only a few agencies. Wherever possible, completed evaluations should be placed in the public domain.

13

Follow-up

Evaluation reports need to be ‘sold’. Donors, field officers and agency staff need to be enthused, excited and convinced that the evaluation report is important and should be read. While selling the report is more the responsibility of the management group than the evaluation team, marketing strategies could be included in negotiated follow-up actions in order to help steering committee members sell the evaluation report within their own organization.

Large system-wide evaluations raise issues which relate to a diverse range of organisations and compliance cannot be compelled. As a consequence, monitoring of follow-up action is important because it provides for a level of accountability which is otherwise missing. A well-resourced and well-structured monitoring process can strongly influence agencies to account for their response (or lack of it) to the evaluation report. The Joint Evaluation Follow-up Monitoring and Facilitation Network (JEFF), established after the Rwanda evaluation provides a potential model for institutionalising post-evaluation monitoring (JEFF 1997).

It would seem that the evaluation process would be more constructive if managers established a mechanism whereby they made formal decisions on the conclusions of an evaluation, explaining where action was to be taken and who

... was responsible. Where no action was deemed appropriate, this would need to be justified. Evaluation departments would then have a useful tool with which to ensure that the evaluations were followed by the necessary action needed to improve policy and practice.

Box 21

Following-up evaluations

A small group, representing the Steering Committee that had managed the Rwanda Evaluation, met regularly in the year following publication of the evaluation to report back to the Steering Committee on action taken on recommendations by the bodies concerned. This maintained interest in the evaluation report, and kept up pressure on concerned bodies to demonstrate that they were taking the report seriously.

WFP have recently formed a Recommendation Tracking Unit, composed of individuals drawn from senior management who meet to check to see what has happened to the recommendations made in evaluation reports. The unit's aim is not to ensure that recommendations have been adopted, but rather that recommendations have been debated at the appropriate level. Recommendations are tracked for 18 months following an evaluation, the evaluation process being seen as something that takes place over 2.5 years. Tracking begins with a discussion of what the key issues are, followed by preliminary contacts with the field, and sometimes a pre-evaluation trip to confirm priorities, track down key informants, start looking for secondary data sources and prepare for primary data collection. It ends with a report on whether recommendations have been considered.

Following the OLS review, the UN agencies that made up the OLS consortium produced a "Plan of Action for Implementation of Recommendations". An initial response to the Review team's recommendations were made, and an action plan drawn up. Every quarter, the consortium monitored progress against the plan of action.

Annex I

Characteristics of complex emergencies and the international system for responding to them

At the risk of oversimplification, the principal characteristics of a ‘typical’ complex emergency and a ‘typical’ response to it by the international community are identified below, the purpose being to draw attention to those areas in which there are likely to be particular challenges for those undertaking evaluations in such environments.

Collapse of state functions

Complex emergencies are frequently both the product of and a factor in the collapse of state functions. The traditional framework for offering humanitarian assistance, designed to address relief in natural disasters, has regularly proven inadequate when no state interlocutors can be identified.

Intra-state rather than inter-state conflict

The majority of conflicts are now intra-state (often termed civil wars) rather than inter-state, though many of them also spread into neighbouring countries, particularly where an ethnic group is found on both sides of a national border. Most intra-state conflicts result from a disaffection by part of the population with the existing structures of governance and authority. In many situations the declared objective of opposition or ‘rebel’ groups is either to overthrow the internationally recognised government or to secede part of the country to establish a new autonomous entity. The degree of recognition or status given to opposition groups or warring factions by international organisations and the international community is often a highly sensitive issue. This context has important implications for the role of international agencies, the legal basis for their operations and the legal rights of affected populations within the affected countries.

Difficulty in differentiating combatants and civilians

In many recent intra-state conflicts it has proven difficult to distinguish between civilians and combatants. Uniforms are often not worn. Combatants may be fed by their families or food procured from the local population either on a voluntary basis or through coercion. In many cases the intermingling of

combatants and civilians is a deliberate policy. In such a situation, humanitarian agencies are often unable to prevent assistance distributed for genuine civilians and vulnerable populations being used by combatants and warring factions.

Violence directed towards civilians and civil structures

In many contemporary conflicts civilians are deliberately attacked and their way of life undermined in order to displace particular social or ethnic groups or, in some cases, to eliminate them. As well as attacks on economic infrastructure (e.g. government buildings, factories, roads and railways), attacks on communities are planned to instill such fear in neighbouring communities and groups of similar social or ethnic background that they are forced to seek refuge elsewhere. In this way territory, farmland and housing may be released for use by the faction responsible for the atrocity. Attacks may also be carried out on targets which play a special role in the cultural identity of particular groups (e.g. places of religious worship and culturally significant buildings).

Fluidity of the situation on the ground

While natural disasters often involve sudden and traumatic events (e.g. earthquake, floods, cyclones) the duration of the events is rarely more than a few days, though it may take months or years for the affected population to resume their pre-disaster livelihoods. However, complex emergencies frequently last several years. In the case of Eritrea the conflict lasted three decades and in the cases of Afghanistan, Mozambique and Angola almost two decades. Within such chronic situations the situation can be highly fluid in particular geographical areas. Fighting may produce civilian as well as military casualties and threatened populations may often flee creating more or less localised displacement crises. Chronic problems are therefore frequently interspersed with situations requiring urgent responses by humanitarian agencies. There is almost never a linear relief-development continuum.

Lack or absence of normal accountability mechanisms

While the degree of functionality and freedom enjoyed by the press and judiciary may be limited in many states not affected by instability or conflict, in most complex emergencies the degree of functionality and freedom is either severely constrained or has been eliminated. Those involved in the conflict and those involved in trying to provide humanitarian assistance therefore operate in a context of absent or severely weakened national accountability mechanisms.

The potential and actual development of war economies

The absence of accountable authorities in parts of or all of a conflict-affected country often results in the development of economic activities normally classified as illegal or semi-legal. Such activities may involve the exploitation and exportation of mineral deposits and natural resources, the cultivation and exportation of narcotics, the laundering of ‘dirty’ money and trading in arms. Often these activities are controlled by, or heavily taxed by, leaders of the warring factions who use the resources either for personal gain or to prosecute the conflict. Because of the international character of the transactions such illegal and semi-legal activity invariably involves the participation of commercial organisations based outside the affected countries in neighbouring countries, industrialised countries or other countries. In some cases such activity is carried out with the knowledge of the authorities. In some complex emergencies such ‘war economies’ have enabled a significant prolongation of the conflict.

The potential for humanitarian assistance to prolong the conflict

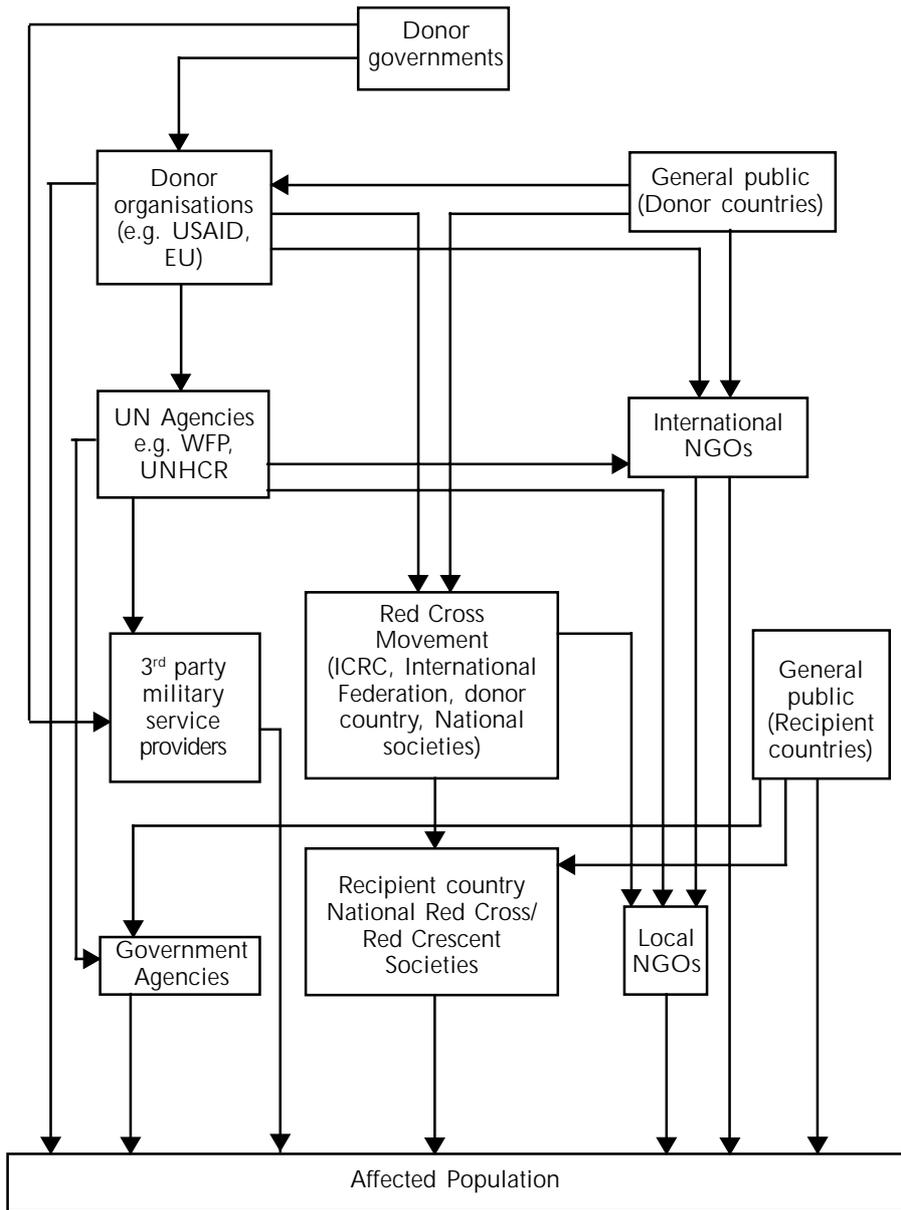
As a resource being provided into areas of ongoing conflicts, humanitarian assistance is potentially capable of being ‘diverted’ from the intended beneficiaries and controlled and taxed in the same way as ‘war economy’ activities. Reliable empirical evidence of the degree to which humanitarian assistance is diverted is often lacking even though such knowledge may be fundamental to understanding the impact of the assistance provided either on the intended target group or in potentially providing warring factions with additional resources.

A multiplicity of actors

The complexity of the international system for responding to complex emergencies cannot be overstressed. Figure 2 (on page 102) conveys some of the complexity by showing the principal routes of resource flows within the system. The organisations commonly involved in the provision of humanitarian assistance include national relief structures (in areas where they are still operating) or relief structures associated with particular factions, local NGOs, UN agencies, international NGOs, the ICRC, the IFRC and the national Red Cross and Red Crescent Society.

In some instances, large bilateral and multilateral donor organisations may establish local field teams with the objective of funding locally prepared projects

Figure 2
Resource flows within the international relief system



and coordinating their overall activities. Increasingly, human rights agencies and monitors are also deployed in ongoing conflicts as are organisations seeking to resolve the conflicts. In those situations where international peacekeeping or peace enforcement forces are deployed either by the UN or by regional bodies (eg. OAU, ECOWAS, CIS), humanitarian activities will run in parallel to, or in concert with, the peacekeeping operations. Such military interventions may involve troop contingents from a variety of countries. In situations where displaced populations cross international borders, refugee agencies will become involved together with the governments of the asylum countries. It is not uncommon for neighbouring states to be involved (either overtly or covertly) in intra-state conflicts in neighbouring countries. States with traditionally strong links with the conflict-affected country or which perceive their strategic interests to be at stake may also play a role in relation to the conflict. A typical complex emergency may therefore involve 8 UN agencies, the Red Cross Movement (ICRC, IFRC and the National Society), 50 or more international and local NGOs involved in the provision of humanitarian assistance, human rights activities and conflict-resolution activities, military contingents operating either under a UN or a regional body commander and agencies controlled by or associated with the warring factions. Their security and humanitarian activities may be supported by 20 or more official donors supplemented by funds raised privately. In addition a number of neighbouring and powerful states may be taking a keen interest in the course of the conflict and possibly seeking to influence the outcome by overt or covert means.

Given that central authority structures are generally weak or absent, coordination of such a multiplicity of actors becomes a major challenge – one that has not yet been fully met by the international humanitarian system. Coordination structures and mechanisms which exist within the UN system and the NGO community are rarely strong and effective. In addition, the wide range of actors involved in the response or with an interest in the outcome of the conflict almost guarantees that their goals will not be shared and may even be at odds with each other (see box overleaf).

An example of differences in the goals of key actors

In managing refugee settlements near the Kenyan/Somali border in 1991, UNHCR had to balance a number of diverse interests and pressures. The camps initially provided a safe haven for those fleeing armed conflict, but they also attracted refugees from Somalia who were not displaced by war, as well as Kenyan nationals wishing to access repatriation benefits. Similarly, the goal of the Government of Kenya – rapid repatriation of the refugee population – was not shared by UN agencies. UNOSOM, for example, was advising against repatriation because of the continuing insecurity in large areas of Somalia.

Source: Kirkby, Kliet & O'Keefe, 1997:15, cited in Apthorpe and Nevile, 1998

Annex II

The logical framework

Given the large number of publications on the LogFrame process, this Annex will limit itself to a brief introduction, highlighting some of the limitations of LFA.

A logical framework is a tool used in the design, planning, implementation, monitoring and evaluation of a project. The log framework (LF), if correctly applied, exposes the logic of how a project is expected to work, by setting out the various levels in the hierarchy of objectives (see Figure 3 on page 106 as an example of a standard LF).

In the first column, there are four rows, which outline the logic of intervention. The goal of the project is set out at the top of a LF, and is a statement of the wider problem that the project will help to resolve. Below the goal, is the purpose of the intervention, namely the direct impact on the project area or target group to be achieved by the project. The outputs are the specifically deliverable results expected from the project to attain the purpose. At the bottom of the LF are the activities, those tasks which need to be done to produce the outputs.

In the second column are the measurable indicators (often called the objectively verifiable indicators (OVIs)) – an operational description of the goal, purpose and outputs in terms of quantitative or qualitative results. The bottom row contains the physical and non-physical inputs needed to carry out the planned activities.

The third column details the sources of verification, which indicate where and in what form information on the achievement of project purpose and results can be found. The fourth column details the assumptions – the external factors outside the direct control of the project that are crucial for the achievement of the activities, outputs, purpose and goal.

The rows represent a ‘hierarchy’ of objectives: the bottom row describes the activities, through which the outputs are to be achieved which, in turn, are to lead to the achievement of the project purpose, which then contributes to the overall goal.

Figure 3

A standard logframe

Objectives	Measurable indicators	Means of verification	Important assumptions
<p>GOAL: <i>Wider problem the project will help to resolve</i></p>	Quantitative ways of measuring or qualitative ways of judging timed achievement of goal	Cost-effective methods and sources to quantify or assess indicators	<p>(Goal to supergoal) External factors necessary to sustain objectives in the long run</p>
<p>PURPOSE <i>The immediate impact on the project area or target group i.e. the change or benefit to be achieved by the project</i></p>	Quantitative ways of measuring or qualitative ways of judging timed achievement of goal	Cost-effective methods and sources to quantify or assess indicators	<p>(Purpose to goal) External conditions necessary if achieved project purpose is to contribute to reaching project goal</p>
<p>OUTPUTS <i>These are the specifically deliverable results expected from the project to attain the purpose</i></p>	Quantitative ways of measuring or qualitative ways of judging timed achievement of goal	Cost-effective methods and sources to quantify or assess indicators	<p>(Outputs to purpose) Factors out of project control which, if present, could restrict progress from outputs to achieving project purpose</p>
<p>ACTIVITIES <i>These are tasks to be done to produce the outputs</i></p>	<p>INPUTS This is a summary of the project budget</p>	Financial outturn report as agreed in grant agreement	<p>(Activity to output) Factors out of project control which, if present, could restrict progress from outputs to achieving project purpose</p>

Source: DFID, 1998

The project purpose is “the key point of reference, the true ‘centre of gravity’, for project management and permitting measurement of the project’s success or failure in terms of sustainable benefits for the target group. The project management is expected to ensure that the project purpose is being achieved” (EU, 1997).

The existence of a LF allows evaluators to see clearly the thinking and the strategy behind the intervention, and to test the internal logic of the intervention. It should also allow them, through the use of the objectively verifiable indicators, to measure project success.

Some agencies have adapted the LF, for example, adding a row for activities, between the rows for inputs and outputs, or removing the column on indicators, in order to reduce the focus on the quantitatively measurable aspects of performance. Whatever the changes, however, the basic concept remains largely the same.

There are a number of reservations about the use of LFs (Gasper, 1997 and Gasper 1998):

- i) the process is too rigid, particularly for humanitarian aid programmes in complex emergencies, where events can unfold very rapidly, completely altering the background and the objectives of a programme;
- ii) there is confusion between the meanings of some of the terms, in particular between purpose and goal and between purpose and outputs;
- iii) it can rarely be assumed that inputs will lead automatically to outputs, and the link between outputs and purpose (and goal) can be very tenuous or difficult to prove. Whether a project approach is logical depends on considerably more than just filling in a LF matrix. It requires solid analysis;
- iv) the assumptions column is often not treated with the seriousness it deserves. This may be because the LF grew up in the context of engineering projects where there was a fairly controlled environment. In a complex emergency, however, the assumptions column could be overwhelmed;
- v) there is an assumption underlying the LF that interventions happen in a relatively closed and manageable world, when the opposite is generally true; and,

- vi) use of the logframe can lead to ‘tunnel vision’ of evaluators who concentrate overly on intended effects.

Gasper goes on to suggest a modified logical framework analysis, one in which assumptions analysis is given special attention. To achieve this:

- i) make the assumptions column (or set of columns) the first. “The conventional LF layout, tends to mean that indicators are defined in advance of adequate causal analysis, and assumptions analysis becomes a belated confidence-bolstering ritual, because, by then, so much effort has been invested in the design concept” (Gasper, 1998); and,
- ii) increase the number of assumptions columns. Add a column for indicators to monitor the key external factors about which assumptions are being made, and a column for comments on assumptions. (Possibly, however, in a complex emergency, there will be so many important external factors that a matrix will not be adequate).

Figure 4 below is an example of a modified logframe, proposed in the paper by Gasper. However, even a modified logframe will need to be used as a flexible tool, and not as an ‘iron frame’ of indicators and targets that hinder innovation and improvisation.

Figure 4

A modified logframe

Sources of info on assns.	Indicators on key external factors	Comments on assns.	assns about key external factors	levels of objectives	indicators for objectives	sources of info. for objectives indicators	Targets (if any) for achievement

Annex III

Questions for evaluators²⁰

It may be useful for evaluators to refer to a checklist of questions commonly asked in evaluations. A list of potential questions is presented below (drawing heavily from Martin, L. 'Planning and Organising Useful Evaluations', UNHCR, January 1998).

Understanding the situation

- How was the initial situation assessed? What access was there to the affected populations? What data was collected, and how was it analysed?
- What would have happened if no assistance had been given?
- Were lives in danger?
- What are the security implications of providing/not providing assistance? Were these considered?
- Are new movements of people/other significant events possible? Have contingency plans been made for this?
- How have beneficiary groups been identified? Using what characterisation? Were the special needs of certain groups (children, female-headed households, disabled, ethnic minorities) considered? Are any groups (unintentionally) favoured over others as a consequence of the way the programme was set-up? Are beneficiaries unfairly assisted in relationship to host/other populations?
- Was the impact of the emergency and the humanitarian programme on host populations and the local environment considered?
- What background research was carried out on the culture and history of beneficiary groups? How were information gaps filled?
- What efforts were made to obtain reliable disaggregated population figures?
- What range of solutions were considered to the problem?

- Were other agencies included in analysis of the problem and the drawing up of response plans? Were response plans ‘holistic’ or very much determined by the mandate of individual organisations?
- What analysis has been undertaken of possible long-term solutions to the problems? – What is the quality of this analysis?
- Were preparedness activities carried-out?

Reviewing the plan

- Have lessons been learnt from earlier programmes?
- Did original plans contemplate exit strategies?
- Were constraints to implementation adequately assessed from the outset?
- Were all critical factors considered, including the political, economic, and cultural factors, as well as the weather and logistical considerations? What subsequent changes took place in the institutional, political and economic context? What impact did these changes have on programmes?
- What major assumptions were made at early stages of the response? Did these turn out to be well founded? If not, in what respect?
- Were there major considerations that should have been foreseen but were not?
- Were the programme objectives sufficiently clear? Were alternative approaches to achieving the objectives considered? Are objectives sufficiently explicit so that their achievement can be assessed and impact determined?
- Are the activities within the mandate of the organisation concerned?
- Did activities correspond to actual needs? Are they consistent with solutions?
- What were the main inputs, outputs and activities?
- Do programme plans take into account longer-term (re-)integration of beneficiaries?
- Do activities discourage initiative and lead to dependence?

- What was the administrative, personnel and logistic back-up like? Did short-comings in these areas have any significant impact on performance?
- Was an assessment made of the links between assistance and protection activities?
- Did staff have the experience and expertise to carry-out the activities envisaged?
- Did headquarters and field staff agree on programme aims, priorities and implementation strategies? Did lack of agreement cause any substantial difficulties?
- Have cost considerations been a part of the planning and design process?

Design

- Does the operational/conceptual model/ logical framework underlying the intervention make sense? Will the activities proposed achieve the desired outputs? Will these outputs lead to the desired changes (impact)? Have reasonable assumptions been made about what may affect the projected outcome? Have important assumptions been left out or ignored?

Involvement of beneficiaries²¹

- Were the interests of beneficiaries adequately taken into account in the design and implementation of the programme? Or were procedures to canvass the opinion of beneficiaries a token gesture, which over-represented the views of the controlling elite, or those who were best able to articulate their views?
- Were beneficiary groups equitably and transparently identified for targeting and were they involved in the negotiation of beneficiary status?
- Was the aid provided sufficient for need at the time and did it relate to livelihood and coping strategies?
- Was aid culturally acceptable and appropriate?
- Did aid meet negotiated targets?
- Was aid delivered in a form that allowed ease of transport for the recipient?

- Was the delivery system transparent?
- Was the delivery of aid targeted directly to those in need, taking into account limitations on their ability or inability to travel?
- Was there a rapid response by agencies to failures in the delivery system?
- Did aid allow people to make autonomous decisions about their own lives?
- Was human dignity maintained in agency dealings with beneficiaries?
- Did agencies actively seek and listen to the beneficiaries' views on aspects of the aid system?
- Were the security and safety of the beneficiaries high priorities?

Assessing the context and linkages

- What are the government's objectives, and how do they differ?
- How can different agendas (whether between agencies, between agencies and government or between government and other national forces) complicate the response? Have measures been taken to mitigate the negative impact of different agendas, or to try and bring those with different agendas together?
- Have national resources been used appropriately?
- Are donors in agreement about programme objectives?
- Was the best-placed organisation (in terms of knowledge of the country, skill profile, capacity, etc.) used to carry out the programme?
- Are activities well coordinated? Are sectoral responsibilities successfully divided between partners?
- Are plans well communicated (to other partners, governments and beneficiaries)?

Analysing Implementation

- Were assumptions accurate?
- Were plans followed? If not, why not?

- Have objectives changed during implementation? Why? On the basis of what analysis or what events?
- Are assistance activities contributing to protection?
- Are lines of authority and reporting clear?
- Are there staff or skill shortages? What has been done about these?
- Did staff receive adequate briefing?
- Did activities match needs?
- Did beneficiaries participate in programme implementation? What was the impact on implementation of their (non-)participation?
- Have there been delays in implementation? How were these dealt with? What was the impact of such delays?
- Has there been coordination and information-sharing with other agencies, governments, donors, beneficiaries? Has this avoided duplication and led to sensible sharing of tasks between actors?
- Was decision-making undertaken in a timely manner and delegated to the appropriate level?
- Have correct financial and administrative procedures been followed?
- Are activities consistent with policies, recognised good practice and guidelines?
- What problems have emerged, for what reasons and how have they been dealt with?
- What have been the strengths and weaknesses of implementation?
- Where could activities have been more cost-effective?
- What training of partners has taken place? What capacity has been left behind?
- Are programme activities regularly monitored and evaluated?

Evaluating results

- Were measures to provide security and protection adequate?
- Were those most in need prioritised for assistance?
- Did activities contribute to solutions? In what way?
- Was social conflict prevented or limited?
- Will activities continue once agencies pull out?
- Were needs overlooked or neglected?
- What is the analysis of the current situation? How are events expected to develop?
- Were commitments honoured? Did the assistance make a difference? What were the most successful activities? And the least successful? Would alternative forms of assistance been more effective?
- What were the long-term impacts of assistance?
- What will happen once agencies pull out?
- How do the achievements compare to plans and objectives?
- Were resources effectively and efficiently used? Do the results justify the costs?
- Was assistance significant or of marginal value to beneficiaries?
- Were objectives relevant and worth pursuing?
- Where is improvement needed?
- What new programme, policies, activities, systems or guidelines are required?
- What lessons can be learned?

Acronyms

ALNAP	Active Learning Network on Accountability and Performance in humanitarian assistance
ANU	Australian National University
DAC	Development Assistance Committee (of the OECD)
DFID	Department for International Development (UK)
DSA	Development Studies Association (UK)
EU	European Union
GPR	Good Practice Review
ICRC	International Committee of the Red Cross and Red Crescent
IDP	Internally displaced person
IDS	Institute of Development Studies (UK)
IFRC	International Federation of the Red Cross and Red Crescent
JEFF	Joint Evaluation Follow-up monitoring and Facilitation network
LF	Logical Framework
MSF	Médecins sans Frontières
NCFS	National Centre for Development Studies (Australia)
NGO	Non-Governmental Organisation
OAU	Organization for African Unity
OCHA	Office for the Coordination of Humanitarian Affairs
ODA	Overseas Development Administration, now DFID
ODI	Overseas Development Institute, London.
OECD/DAC	Organization for Economic Cooperation and Development/ Development Assistance Committee
OLS	Operation Lifeline Sudan
OVI	Objectively verifiable indicator
PRA	Participatory Rural Appraisal
REMAPP	Research, Evaluation, Monitoring, Appraisal, Planning and Performance
RRA	Rapid Rural Appraisal
SCF	Save the Children Fund
SCHR	Steering Committee for Humanitarian Response
ToR	Terms of reference
UNDP	United Nations Development Program
UNHCR	United Nations High Commissioner for Refugees
WFP	World Food Programme
WHO	World Health Organization

Endnotes

1. At the 30th meeting of the DAC Working Party on Aid Evaluation in May 1998, a draft of a related paper, entitled “Good Practice in the Evaluation of Humanitarian Assistance Programmes in response to Complex Emergencies: Guidance for Evaluation Managers ” was submitted by DANIDA as Room Document No.3. This GPR, while intended for a slightly different audience, contains the main findings and conclusions of this earlier paper.
2. ALNAP’s full membership includes representatives from a broad selection of bilateral donor organisations and governments, multilateral donor organisations, UN agencies and departments, NGOs and NGO umbrella organisations, the International Movement of the Red Cross and the Red Crescent and consultants and academics.
3. Development Assistance Committee, ‘Principles for Evaluation of Development Assistance’, Organisation for Economic Co-operation and Development, Paris, 1991
4. As per the terminology established by the Expert Group on Aid Evaluation established by DAC in 1982. See Berlage and Stokke: 1992
5. Development Assistance Committee, ‘Principles for Evaluation of Development Assistance’, Organisation for Economic Co-operation and Development, Paris, 1991
6. See Van Brabant (October, 1997) for a more detailed discussion of institutional and organisational learning, as well as Minear (April, 1998).
7. Personal communication with Phil O’Keefe, ETC UK
8. Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda, 1996, ‘The International Response to Conflict and Genocide: Lessons from the Rwanda Experience’, Copenhagen.
9. This section draws heavily upon both personal discussion with Raymond Apthorpe and a report on the findings of a workshop held, with AusAID support, at the National Centre for Development Studies (NCDS), Australian

National University (ANU), in Canberra, from 13th to 16th March 1998: 'Managing Emergency Humanitarian Aid Evaluation: Lessons from Experience', by Apthorpe and Nevile, April 1998. Thanks are extended to the authors of this report, soon to be published in Apthorpe, R. (1998) 'Towards evaluating humanitarian aid', NCDS Occasional Issues Paper, ANU, Canberra.

10. It is important to note, however, that the use of the term audit is currently changing, as concepts such as 'social audit' gain currency.
11. A Logical Framework Analysis is a tool used in the design, planning, implementation, monitoring and evaluation of a project. It exposes the logic of how a project is expected to work, by setting out the various levels in a hierarchy of objectives. See Annex II on page 105.
12. The People in Aid Code of Best Practice was first published by the Overseas Development Institute as a RRN Network Paper. It has since been reproduced by the People in Aid Project and copies are available from People in Aid, BRCS, 9 Grosvenor Crescent, London, SW1 7EL, UK. Tel./Fax +44 (0)171 2350895.
13. Care will be needed, however, where on-going conflict means that impartiality is unlikely to be possible.
14. In the same way that it has become the practice of offering counselling to returning relief personnel it may be appropriate to offer counselling to returning evaluators.
15. Indeed, Harvey (IDS, 1997), in a review of the use of PRA in emergencies, found little evidence that such techniques are being widely used, whether for evaluation or other management purposes.
16. This section draws heavily on Hallam, A. (1996)
17. A project can be cost-efficient, without necessarily being effective. For example, a school could be efficiently built, but a lack of teachers mean that the objective of increasing school attendance will not be achieved.
18. The purchase and transport of food aid alone can absorb 40% of a typical relief programme budget.

19. Evaluators might like to refer to other Good Practice Reviews in the RRN series – on the provision of water and sanitation in emergencies, on emergency supplementary feeding programmes, on general food distribution in emergencies, on seed provision during and after emergencies, on the counting and identification of beneficiary populations in emergency operations and on temporary human settlement planning for displaced populations in emergencies. Such publications offer as much to the evaluator of humanitarian assistance programmes as books on evaluation methodology.
20. Adapted from Martin, L. ‘Planning and Organising Useful Evaluations’, UNHCR, January 1998
21. This section draws from personal correspondence from John Kirkby and Phil O’Keefe of ETC UK

References

- Apthorpe, R (1996) *'Protracted Emergency Humanitarian Relief Food Aid: Towards Productive Relief'* Programme Policy Evaluation of the 1990-95 period of WFP-assisted refugees and displaced persons operations in Liberia, Sierra Leone and Cote d'Ivoire, Rome.
- Apthorpe, R and Gasper, D. (1996), *'Arguing Development Policy: Frames and Discourses'*, Frank Cass and Co., London.
- Apthorpe, R. and Nevile, A. (1998) 'Managing Emergency Humanitarian Aid Evaluation: Lessons from Experience' A report on the main findings of a workshop held with AusAID support at the National Centre for Development Studies (NCDS), Australian National University 13-16 March 1998, forthcoming in Apthorpe, R. (1998) 'Towards evaluating humanitarian aid', *NCDS Occasional Issues Paper*, Australian National University, Canberra.
- Berlage, L. and Stokke O. (1992) *'Evaluating Development Assistance: Approaches and Methods'*, Frank Cass, London
- Borton, J., Brusset, E., Hallam, A. (1996), "Joint Evaluation of Emergency Assistance to Rwanda: Humanitarian Aid and Effects", ODI, London.
- Borton, J and Macrae, J. (1998) 'Evaluation Synthesis of Emergency Aid' *EV 613* Department for International Development, London.
- Chambers, R. (1992) 'Rural Appraisal: Rapid, Relaxed and Participatory', *Institute of Development Studies Discussion Paper No. 311*. Brighton: IDS
- 'Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief', (1994) *Relief and Rehabilitation Network Paper 4*. London: Overseas Development Institute.
- Cracknell, B. (1996), "Evaluating Development Aid: Strengths and Weaknesses" in *Evaluation*, Vol 2(1): 23-33. London: SAGE Publications.
- Darcy, James (1997), "Human rights and international legal standards: what do relief workers need to know?", *Relief and Rehabilitation Network Paper 19*. London: Overseas Development Institute.

Donini, A, Dudley, E. & Ockwell, R (1996) 'Afghanistan: Coordination in a Fragmented State' a *Lessons Learned Report* prepared for the United Nations Department of Humanitarian Affairs. Geneva: UNDHA.

Fishman, D. (1992), "Postmodernism comes to Program Evaluation: A Critical Review of Guba and Lincoln's Fourth Generation Evaluation", *Evaluation and Program Planning*, Vol 15: 263-270.

Gasper, D. (1997) 'Logical Frameworks: A Critical Assessment', Working Paper no. 278, The Hague: Institute of Social Studies

Gasper, D. (March 1998), 'Notes on the Logical Framework Approach and the Planning and Evaluation of (Complex) Emergency Relief', The Hague: Institute of Social Studies

Griekspoor A., Sondorp E. and Vos T., (Date unknown) 'Introducing cost-effectiveness analysis of humanitarian relief interventions: a case study of a visceral leishmaniasis treatment program in the Sudan. Presently under submission to HP&P.

Hallam, A. (1996) 'Cost Effective Analysis: A Useful Tool for the Assessment and Evaluation of Relief Operations?' *Relief and Rehabilitation Network Paper 15*, April 1996. London: ODI.

Hallam and Borton, (1998), 'Good Practice in the Evaluation of Humanitarian Assistance Programmes in Response to Complex Emergencies: Guidance Material for Evaluation Managers': 'Product B' of the Study to Identify and Disseminate Best Practice in the Evaluation of Humanitarian Assistance Programme. Draft tabled for consideration by the DAC Working Party on Aid Evaluation at its meeting of 26-27 May 1998.

Harvey, P. (1997) *PRA and Participatory Approaches in Emergency Situations: A Review of the Literature and ACTIONAID's Experience*. London: ACTIONAID.

Hulme, D. (1997) 'Impact Assessment Methodologies for Microfinance: A Review', *Paper Prepared for the Virtual Meeting of the CGAP Working Group on Impact Assessment Methodologies*, April 17-19, 1997. Manchester: IDPM.

Jaspars, S., Young, H., Shuria, H., Ogolla, L., Kisopia, P., Munyes, J. & Kapoko, A. (1997) *People on the Edge: An Evaluation of Oxfam's Emergency Intervention in Turkana, March-August 1996*. Oxford: OXFAM.

- JEFF 1997. 'The Joint Evaluation of Emergency Assistance to Rwanda: A Review of Follow-up and Impact Fifteen Months After Publication' Sida, Stockholm and ODI, London
- Karim, A, et al. (1996) 'Operation Lifeline Sudan: A Review' DHA-Geneva/ University of Birmingham.
- Kirkby, J., Kliet, T. & O'Keefe, P. (1997), 'The Complex Emergency in Somalia: UNHCR's Cross Border Cross Mandate Operation. Netherlands Ministry of Foreign Affairs, The Hague.
- Kruse, S., Kyllonen, T., Ojanpera, S., Riddell, R., Vielajus, J-L., with Bebbington, A., Humphreys, D. & Mansfield, D., (May 1997) "Searching for Impact and Methods: NGO Evaluation Synthesis Study: A Report prepared for the OECD/DAC Expert group on Evaluation". Paris: OECD.
- Lindhahl, C. (March 1998), 'The Management of Disaster Relief Evaluations: Lessons from a Sida evaluation of the complex emergency in Cambodia', Solna, Sweden (email: mpi@swipnet.se)
- Macrae, J. (1997), "Purity or Political Engagement?: Issues in food and health security interventions in complex political emergencies", (Draft). London: ODI.
- Macrae, J. (1997), "Aid, Geopolitics and Principles", *Notes for a presentation to the DHA-UNDP workshop on building bridges between relief and development*. London: Overseas Development Institute.
- Maxwell, S. (1997) 'The Use of Matrix Scoring to Identify Systemic Issues in Country-Programme Evaluation', *Development in Practice*, Vol. 7:4 pp.408-415.
- Martin, L. (Jan 1998) 'Planning and Organising Useful Evaluations', UNHCR, January.
- Minear, L. (1994) 'The International Relief System: A Critical Review' Paper presented to the Parallel National Intelligence Estimate on Global Humanitarian Emergencies, Meridian International Centre, Washington DC. September 22
- Minear, L. (April 1998), 'Learning to Learn', Discussion Paper prepared for a seminar on Lessons Learned in Humanitarian Coordination, OCHA and Swedish Ministry of Foreign Affairs, Stockholm.

- · · Mitchell, J. (1997) 'The Listening Legacy: Challenges for Participatory Approaches', *Paper presented at Mitigating the Millennium, 9 October 1996*. London: British Red Cross
- · · Neefjes, K. and David, R. (Oct 1996), 'A participatory review of the Ikafe refugee programme', Oxfam, Oxford.
- · · OECD Development Assistance Committee (1991), *Principles for Evaluation of Development Assistance*, Paris: OECD.
- · · Perea, W. & Paquet, C. (1996) 'Use of Epidemiological Tools for Evaluating Impact in Health Interventions' in MSF Holland's *Mini Symposium on Evaluations and Impact Studies of Humanitarian Relief Interventions*, final report of the symposium held in Amsterdam, the Netherlands, November 29 and 29, 1996. Amsterdam: MSF Holland.
- · · Randel, J. (1994), "Aid, the Military and Humanitarian Assistance: An Attempt to Identify Recent Trends", *Journal of International Development*, 6 (3):329-342.
- · · Ricoeur, P. (1979), 'The Model of the Text: Meaningful Action Considered as a Text', in P. Rabinow and W. Sullivan (eds.), *Interpretative Social Science: A Reader*, Berkely, CA: University of California Press
- · · Slim, H. & Mitchell, J. (1990) *The Application of RAP and RRA Techniques in Emergency Relief Programmes. Mimeo*. Cornwall: Slim and Mitchell Rural Evaluations.
- · · Slim, H. & Mitchell, J. (1990) *The Application of RAP and RRA Techniques*. Cornwall: Rural Evaluations.
- · · Slim, H. & Mitchell, J. (1991) *The Application of RAP and RRA Techniques in Emergency Relief Programmes*. Cornwall: Rural Evaluations.
- · · Slim, H. & Mitchell, J. (1997) 'The Application of RAP and RRA Techniques in Emergency Relief Programmes' in Scrimshaw, N. S. & Gleason, G. R. (Eds) (1997) *Rapid Assessment Procedures: Qualitative Methodologies for Planning and Evaluation of Health Related Programmes*. Boston: INFDC.
- · · Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda, (1996), 'The International Response to Conflict and Genocide: Lessons from the Rwanda Experience', Copenhagen.

UN Draft Strategic Framework for Afghanistan: Advanced Unedited Version, 1998

Van Brabant, K. (Oct, 1997), 'Organisational and Institutional Learning in the Humanitarian Sector: Opening the Dialogue', Discussion Paper for the Active Learning Network on Accountability and Performance in Humanitarian Assistance, ODI, London.

Webb, P. (1996), "Enabling Withdrawal From Emergency Aid", *Paper presented to a Workshop of Natural Resource Advisers*, Overseas Development Administration, Chichester, July.

White, L. (1994), 'Policy Analysis as Discourse', *Journal of Policy Analysis and Management*, Vol. 13, No. 3, pp. 506-25

Wood, A. (March 98), 'Evaluating Emergency Relief Operations: Notes on personal experience of relief evaluation in Ethiopia, Rwanda and Bosnia, for Norad, Sida and the Salvation Army', Huddersfield (a.p.wood@hud.ac.uk).

Further reading

Anderson, M. B. (1996) *Do No Harm: Supporting Local Capacities for Peace Through Aid*. Cambridge, USA: The Collaborative for Development Action, Inc.

Audit Commission (1997) *Feedback '96: Results of the Commission's Consultation on the Local Authority Performance Indicators for 1997/98*. London: Audit Commission.

Borton, J. (1997) 'Synthesis Study of Emergency Aid Evaluations: Supplementary Study on Evaluation Work Undertaken by the European Commission'. London: ODI.

Borton, J. (1997), "Emergency Aid: From Relief Delivery to Mainstreaming Instability", *mimeo*. London: ODI.

Borton, J. (1996), "A Framework for Considering Humanitarian Aid Evaluation", in the Final Report of the MSF Mini Symposium on "Evaluations and Impact Studies of Humanitarian Relief Interventions", held in Amsterdam on 28 and 29 November, 1996. MSF, Amsterdam.

Borton, J. (1994) "Assessing the Impact and Effectiveness of Relief Assistance", *Report prepared for the 1995 World Disasters Report*.

Borton, J. & Macrae, J. (April 1997), "Evaluation Synthesis of Emergency Aid" *Report prepared for the Evaluation Department, ODA*. London: Overseas Development Institute.

Bryce, Jennifer (Date unknown), "Evaluations of Relief Interventions: What to Evaluate", in the Final Report of the MSF Mini Symposium on "Evaluations and Impact Studies of Humanitarian Relief Interventions", held in Amsterdam on 28 and 29 November, 1996. MSF, Amsterdam.

Commission of the European Communities (February 1993), "Project Cycle Management: Integrated Approach and Logical Framework", a *Manual of the Evaluation Unit*, Brussels. Brussels: EC.

Court of Auditors Special Report No 2/97, concerning humanitarian aid from the European Union between 1992 and 1995 together with the Commission's replies, Luxembourg (for further details see <http://europa.eu.int>).

de Veer, T. (1996) *Assessment of Experiences, Attitudes, Beliefs, Developments and Ideas Among Implementing, Donor and Support Agencies with Respect to Monitoring and Evaluation of Water and Sanitation Programmes in Camps for Refugees Or Displaced Persons*. Leiden: Emergency and Humanitarian Aid Bureau of the Netherlands Ministry of Foreign Affairs.

ECHO *Operational Manual for the Evaluation of Humanitarian Aid*, available from ECHO, Brussels (Fax: 32 2 291 11 73).

Edwards, M. (Date unknown), "International development NGOs: legitimacy, accountability, regulation and roles", *discussion paper for the Commission of the Voluntary Sector and the British Overseas Aid Group*.

FAO (1997) *Programme Evaluation Report 1996-97*. Rome: FAO.

Garner, P. (1997), "Do Objectives of health-aid programmes impair their effectiveness?", *The Lancet*, Vol. 349, March 8, 1997, London.

Gosling, L. And Edward, M. (1995), 'Toolkits: A Practical Guide to Assessment, Monitoring, Review and Evaluation', Save the Children Fund, London.

IDS/IDR (1996) 'An Evaluation of Development Cooperation Between The European Union And Ethiopia, 1976-1994' *Sectoral Report: Emergency and Food Aid*. Brighton: Institute of Development Studies; Addis Ababa: Institute of Development Research.

JEFF (1997) *The Joint Evaluation of Emergency Assistance to Rwanda: A Review of Follow-up and Impact Fifteen Months After Publication*, a Report by the Joint Evaluation Follow-up Monitoring and Facilitation Network (JEFF) in conformity with a decision by the Joint Evaluation Steering Committee at its final meeting on 14 February 1997 in Copenhagen. Copies available from ODI: h.awan@odi.org.uk, fax +44 171 393 1699.

Leather, C. (1997) *Assessing Impact: possibilities and constraints in the evaluation of humanitarian assistance*. To obtain a copy contact: c.leather@ich.ucl.ac.uk

Mansfield, D. (1996) *Evaluation: Tried and Tested? A Review of Save the Children Evaluation Reports*. London: Save the Children Fund.

Martin, P. & Quinney, S. (1994), *Participatory Evaluation Project SOS Sahel/UNSO*. London: SOS Sahel.

Minear, L. (1994), *The International Relief System: A Critical Review*, available from Thomas J. Watson Jr. Institute for International Studies (<http://www.brown.edu>).

MSF Holland Monitoring and Evaluation Unit (1997) *Formulating Results for Planning, Monitoring and Evaluation: A Practical Approach*. Amsterdam: MSF.

MSF Holland Monitoring and Evaluation Unit (1997) *Central Monitoring: An Introduction and Tools*. Amsterdam: MSF.

MSF Holland (1996) *Health Systems Research Newsletter/Journal*, September 1996.

MSF Holland (1996) 'Mini Symposium on Evaluation and Impact Studies of Humanitarian Relief Interventions' *Final Report of the Symposium held in Amsterdam, Netherlands, November 28 and 29, 1996*. Amsterdam: MSF.

NRI (1996) *Participatory Rural Appraisal: A Manual on Issues, Principles, and Tools*, Report on Workshop conducted by SSD, Autumn 1996. Kent: NRI.

- OECD/DAC (1997) *DAC Guidelines on Conflict, Peace and Development Cooperation*. Paris: OECD.
- Operation Lifeline Sudan (1996) *A Review, July 1996*. New York: UNICEF.
- Overseas Development Administration (now DFID) (1996), *A Guide to Appraisal, Design, Monitoring, Management and Impact Assessment of Health and Population Projects*. London: Health and Population Division, DFID.
- OXFAM (1997) 'External Trends Seminars', *report from the Seminar on Conflict held at Oxfam House on 29 January 1997*. Oxford: OXFAM.
- Quinn Paton, M. (1990) *Qualitative Evaluation and Research Methods*. London: SAGE.
- Riddell, R. C. (1997) *Evaluating NGO Development Interventions*. London: ODI.
- Royal Ministry of Foreign Affairs, Norway (1993) *Evaluation of Development Assistance: Handbook for Evaluators and Managers*. Oslo: Norwegian Ministry of Foreign Affairs.
- Shaw, I (1995) 'The Quality of Mercy: The Management of Quality in the Personal Social Services' in Kirkpatrick, I. & Martinez Lucio, M. *The Politics of Quality in the Public Sector: the Management of Change*. London: Routledge.
- Shoham, J. (1994) 'Emergency Supplementary Feeding Programmes' *Relief and Rehabilitation Network Good Practice Review 2*. London: Overseas Development Institute.
- Smith, P. (1993) 'Outcome Related Performance Indicators and Organisational Control in the Public Sector', *British Journal of Management*, Vol 4:135-151.
- Sørbo, G. M., Macrae, J. & Wohlgemuth, L. (1997) *NGOs in Conflict: An Evaluation of International Alert*. Bergen: CHR Michelsen Institute.
- UNICEF (1990) 'A conceptual framework showing the causes of malnutrition' in Jaspars, S. and Young, H. 'General Food Distribution in Emergencies: from Nutritional Needs to Political Priorities', *Relief and Rehabilitation Good Practice Review 3*, December 1995.

Valdez, J., Bamberger, M. (Eds) (1994) *Monitoring and Evaluating Social Programs in Developing Countries*, a Handbook for Policymakers, Managers, and Researchers Washington DC: The World Bank.

Warner, M. (1997) "Consensus" Participation: an example for protected areas planning', *Public Administration and Development*, Vol. 17: 413-432. London: Overseas Development Institute.

RRN

Background

The Relief and Rehabilitation Network was conceived in 1993 and launched in 1994 as a mechanism for professional information exchange in the expanding field of humanitarian aid. The need for such a mechanism was identified in the course of research undertaken by the Overseas Development Institute (ODI) on the changing role of NGOs in relief and rehabilitation operations, and was developed in consultation with other Networks operated within ODI. Since April 1994, the RRN has produced publications in three different formats, in French and English: Good Practice Reviews, Network Papers and Newsletters. The RRN is now in its second three year phase (1996-1999), supported by four new donors — DANIDA, SIDA (Sweden), the Department of Foreign Affairs, Ireland and the Department for International Development (UK). Over the three year phase, the RRN will seek to expand its reach and relevance amongst humanitarian agency personnel and to further promote good practice.

Objective

To improve aid policy and practice as it is applied in complex political emergencies.

Purpose

To contribute to individual and institutional learning by encouraging the exchange and dissemination of information relevant to the professional development of those engaged in the provision of humanitarian assistance.

Target audience

Individuals actively engaged in the provision of humanitarian assistance at national and internal, field-based and head office level in the 'North' and 'South'.

This *RRN Good Practice Review* was financed by RRN Phase II donors: