

## Foreword

The third publication in the *ALNAP Annual Review* series echoes many of the main findings from the previous two editions. Most importantly, this year's evaluation synthesis, based on the conclusions and recommendations of 55 evaluative reports, has found that humanitarian action is continuing to meet its fundamental objectives – saving human life and providing basic needs. This is positive news for all of us and provides good grounds for optimism.

However, there is no room for complacency. There are limitations on what we do well and the *Annual Review* has identified a number of additional weaknesses. When evaluated against short-term targets, humanitarian action performs well, but when compared to wider objectives, it almost invariably falls short. The patient may have survived but the health care system is poor; food aid may have staved off starvation but food availability continues to decline; the water system in the camp has helped reduce mortality but spare parts for maintenance are not available.

There may be no easy answers to these shortcomings, but it is up to all of us to address them within the operational frameworks of each of our agencies and organisations. Here the *ALNAP Annual Review* is of great value as it can provide a common basis for developing shared agendas and approaches.

Through the *Annual Review*, ALNAP continues to highlight strengths and weaknesses in the evaluation of humanitarian action through its meta-evaluation quality proforma. I am delighted that, this year, ALNAP is taking a more proactive approach to disseminating these findings by entering into dialogue with agencies whose reports are well represented in the ALNAP Evaluative Reports Database.

Last year's *Annual Review* provided an analysis of key constraints to learning, including the fact that monitoring in humanitarian action was not being utilised in a manner that would optimise its potential for learning. This has been followed up in this year's themed chapter which provides an assessment of current monitoring practice in the sector; it also sets out some practical suggestions on how to strengthen both monitoring and learning, by assessing results and impacts and examining social processes at field level. Monitoring has long been the 'poor cousin' to evaluation and I hope this study reminds us all of its true potential.

I believe that the third *ALNAP Annual Review* demonstrates that the *Annual Review Series* has now come of age. I would like to thank all ALNAP Members who contributed to this edition, either by supplying evaluative reports or through direct contributions, and to encourage all ALNAP Members to use the findings for working towards improved performance in the future.

**Anita L. Menghetti**

Chair of ALNAP (NGO/IO Donor Coordinator, USAID/OFDA)

## Chapter 1

## Introduction

## Purpose and Scope of the ALNAP Annual Review

1.1

The *ALNAP Annual Review* series aims to advance understanding and practice in terms of how the quality of humanitarian action can be improved. It provides a snapshot of current trends in humanitarian action through a synthesis of evaluation findings, as well as an opportunity for more critical reflection on an area of particular concern – this year, the topic of monitoring is addressed. It also provides a platform for sharing lessons, identifying common approaches, and building consensus on directions for improving learning and accountability.

In addition to those working within the humanitarian sector, the *ALNAP Annual Review* series offers valuable insights to those involved in observing and commenting on the sector, for example, journalists, researchers, educationalists, and parliamentary and congressional committee members. Drawing on ALNAP's Evaluative Resources Database, the three years of the *Annual Review* have now covered 145 independent evaluation reports and 20 synthesis reports, perhaps the most exhaustive analysis of evaluation of humanitarian action (EHA) yet accomplished.

### Box 1.1 Purpose and Scope of the *ALNAP Annual Review*

The *ALNAP Annual Review* has three main objectives:

- 1 To provide the humanitarian sector with the means to reflect annually on its performance, and identify generic strengths and weaknesses through a synthesis of the principal findings, conclusions and recommendations of EHA made available to ALNAP during the preceding year.
- 2 To address each year a central theme of common and current concern to those within the sector.
- 3 To monitor and assess the quality of EHA by highlighting good and poor practice through a meta-evaluation of evaluations received the preceding year – a key learning and accountability tool for the humanitarian sector.

The *ALNAP Annual Review* series complements other annual publications focusing on the humanitarian sector, such as the *World Disasters Report* (IFRC) and the *World Vulnerability Report* (UNDP).

## The Emergence of ALNAP and the Accountability Agenda

1.2

As many readers will know, ALNAP emerged from what was perceived as a crisis of accountability in humanitarian action in the mid-1990s. Its mandate was to provide a sectorwide forum, owned by all and dominated by none, as a means of discussing concerns about learning, accountability and quality. At the time ALNAP found itself as part of an emerging group of initiatives that intended to place accountability firmly on the humanitarian agenda (see Table 1.1). However, while a consensus emerged that ‘something needed to be done’ – as humanitarian agencies realised the need to demonstrate they were serious about acting in an accountable manner to beneficiaries, donors and other stakeholders – it was less clear how to do this.

**Table 1.1** Accountability Initiatives in the Humanitarian Sector

| Project Name<br>How to find out more   | Project Objective   | Participants                               |
|--|---|--|
| <b>The Red Cross/<br/>Crescent Code of<br/>Conduct</b> (1994)<br><a href="http://www.ifrc.org">www.ifrc.org</a>                | A voluntary code seeking to guard standards of behaviour in humanitarian action.                              | 207 agency signatories                     |
| <b>People in Aid</b><br>(1996)<br><a href="http://www.peopleinaid.org.uk">www.peopleinaid.org.uk</a>                           | To improve support and management of field staff and volunteers through adherence to a code of best practice. | 11 NGOs/<br>networks                       |
| <b>The Sphere<br/>Project</b> (1997)<br><a href="http://www.sphereproject.org">www.sphereproject.org</a>                       | To develop minimum technical standards and a humanitarian charter for disaster response.                      | A coalition of European and US NGOs        |
| <b>Humanitarian<br/>Accountability<br/>Project (HAP)</b><br>(2000)<br><a href="http://www.hapgeneva.org">www.hapgeneva.org</a> | To find a means of strengthening accountability towards those directly affected by disasters.                 | 13 humanitarian organisations and networks |

*Adapted from Hilhorst (2002).*

ALNAP initially found a niche that stemmed directly from the discussions in the mid-1990s about how to strengthen evaluation practice. The Joint Evaluation of Emergency Assistance to Rwanda (JEEAR) had already demonstrated how evaluation could have a radical affect on the sector, and later the Organisation for Economic Development Assistance (OECD) and the Development Assistance Committee (DAC) Working Group on Evaluation looked at ways to more effectively institutionalise EHA (OECD-DAC, 2001). Since then, ALNAP has been involved in the search for ways in which to combine accountability and learning in evaluation design and practice. This is one of the themes of this *Annual Review*. In these early years more managerial interventions, such as monitoring, received less attention, as it was assumed that an understanding of the problems of past operations would provide the basis and impetus for systemic reform.

## Learning and Monitoring and Evaluation 1.3

The assumption that systemic reform would occur did not come to pass, and over the years it has been found that evaluation needs to be complemented by other approaches to raise accountability and learning during humanitarian operations. Monitoring is gradually being recognised as a key activity in achieving the joint objectives of learning and accountability, and it is partly as a result of this process, in addition to reflection on the findings from the two previous *Annual Reviews*, that ALNAP decided to focus on monitoring this year. Indeed, the topic of monitoring is in many respects a natural progression of the dialogue that ALNAP has developed in the humanitarian community over the past three years.

The first *Annual Review* in 2001 traced the development of humanitarian evaluation and demonstrated that evaluation has become an integral part of how the sector operates. In addition to documenting the considerable achievements that had occurred in the past 10 years, it also analysed the weaknesses. On the plus side, evaluations were found to expose organisations, teams and individuals to critical appraisal and provided a means of assessing when and where problems were addressed. However, there was a lack of clarity as to whether the objective of evaluation was to provide a practical tool for supporting organisational learning or whether EHA was merely intended to create greater upward accountability to

donors. There was a tendency to attempt to combine the two objectives, which in many instances led to lack of attention to the learning agenda (ALNAP, 2001).

Moreover, in 2001 ALNAP commissioned a review of how agencies followed up on the recommendations made in humanitarian evaluations (van de Putte, 2001). The findings showed that recommendations were rarely linked to learning processes and that practical follow-up to evaluations – especially when evaluation was designed as a stand-alone activity – was poor. This of course is not unique to EHA but is endemic in the evaluation field as a whole (Patton, 1997). Too many evaluations are left to collect dust on agency shelves. The need to better integrate evaluation into overall project management cycles and internal policy debates could not have been clearer. This, in turn, required backing from senior management and integration of the planning of evaluation into ongoing management frameworks, such as monitoring.

Last year's *Annual Review* put this in a broader context by providing an analysis of some of the key constraints to learning in the humanitarian sector. This focus on learning was in line with the movement over the last few years from a focus on evaluation methods to a focus on how knowledge produced during evaluation processes can be used to effect change (Patton, 2001). Some worrying observations were made as to the limited impact that evaluation findings and recommendations were having on field practice itself; this corroborated findings from the follow-up study. In other words, why invest so much in evaluation if lessons from past experience are not being learned? One of the main conclusions from last year's *Review* was that, in spite of some improvements in the quality of evaluations, there had been little impact on learning and a corresponding absence of any significant change in humanitarian action (ALNAP, 2002).

Constraints identified in the humanitarian sector included the following:

- Incentive structures in agencies that promote defensive behaviour and a culture of blame.
- Short-term funding mechanisms that militate against a learning environment for field staff.
- Very high rates of staff turnover within ongoing programmes and between programmes.
- Lack of clarity as to intervention objectives and desired outcomes.
- Training provision not properly linked to learning processes.
- Poorly developed mechanisms for cross-organisational learning.

As a result of this analysis an agenda for change and action was developed and presented in the concluding chapter of *Annual Review 2002*. The importance of this agenda has not diminished. Key constraints facing this reform will require much concerted work over a considerable period of time. In this and next year's *Annual Reviews*, ALNAP will take a closer look at the prevailing structures that impede agencies' abilities to learn and will explore how different tools can be used to address these dysfunctions. Specifically, it will ask whether monitoring can become more than a data collection exercise (as is often the case) to instead become a valuable vehicle for 'learning while doing'.

### Monitoring and Learning in Real-time

#### 1.3.1

The experiences of the past few years both within the humanitarian sector and the wider evaluation field have thus contributed to a desire in ALNAP and the humanitarian community to look beyond evaluation to see how other approaches may contribute to learning. Partly as a consequence we are now seeing an increase in new kinds of activities such as Real-time Evaluations (RTE), strategic review and self-evaluation (see Chapter 2). These aim to make up for deficiencies in traditional ex-post evaluation and are intended to provide a timely, rapid review of a particular response so that findings can be used to feed into ongoing decision making processes.

As demonstrated in Chapter 2, UNHCR started the RTE trend, closely followed by WFP, ALNAP's Learning Support Office (Malawi LSO) and the Humanitarian Accountability Project (HAP). The UK Disaster Emergency Committee (DEC) has also carried out some 'monitoring missions' that closely resemble RTE.

Alongside RTE other 'real-time' monitoring practices have increasingly been acknowledged as potentially important though often neglected topics. This is true for both of the main types of monitoring identified in Chapter 2 – situation/context monitoring, and performance/impact monitoring. The renewed interest in these activities is motivated by the concern that ex-post evaluations often come too late to affect the operations they assess and, given the weak institutional memories in many organisations, too early to influence the next operation. Ex-post evaluations will in some cases support learning, in particular where there are mechanisms for the integration of evaluation findings. However, the structure of many agencies militates against such integration: in the UN system and donors, for example,



evaluation offices are often kept separate from programming branches and report directly to the Executive Director's Office – in order to ensure 'independence'. This practice may, ironically, ensure lack of integration of findings into the programming branches.

Monitoring, on the other hand, is an ongoing agency function that is generally integrated into everyday programming. And as opposed to ex-post evaluation, monitoring may be able to offer a 'short action learning loop' (ALNAP, 2002). The importance of this cannot be underestimated given that information from monitoring can potentially enable mid-course corrections of programmes. It could also provide information for humanitarian field workers which could potentially empower them to make better judgements during the course of their immediate work. In this respect one area where monitoring has a distinct advantage is its potential to examine social process, such as why interventions are or are not working, who is benefiting, and why.

But monitoring will require more attention if this is going to happen. Last year's *Annual Review* indicated that the potential benefit of monitoring for learning is not being fully exploited, and conclusions from analysis of the 165 evaluation reports over the three years of the *Annual Review* strongly support the notion that the current quality of monitoring is poor. For these reasons it was decided that the themed chapter in this year's *Review* would directly address the question of whether robust monitoring systems are really able to make up for the deficiencies identified in the two previous *Reviews*. Can monitoring 'fill the gap' by playing a crucial role in both broad sectoral learning and in helping field workers make critical judgements in the midst of humanitarian crises?

## Learning and Downward Accountability: Is Monitoring a Missing Link? 1.4

While evaluation remains an important tool for upward accountability – to donors, headquarters and auditors – it is clear that there are still prevailing learning and downward accountability gaps, the latter especially to the primary stakeholders. Edwards and Hulme (1995) comment that accountability is generally interpreted as

‘the means by which ... organisations report (upwards) to a recognised authority ... and are held responsible for their actions, with insufficient attention directed “downwards” to the views of the intended beneficiaries.’<sup>1</sup> Yet downward accountability has both a practical and an ethical dimension. Practically, improved downward accountability will support closer consultation and participation of affected people in the design and implementation of interventions. This means interventions are more likely to reflect genuine needs and priorities, and achieve optimum impact. Ethically, downward accountability is embedded in the values and principles central to humanitarian action, notably the Red Cross/Crescent Code of Conduct and the Humanitarian Charter. Thus a commitment to downward accountability is part of the living value system that underpins humanitarian action itself.

Initial findings from the ALNAP-commissioned Global Study on the Consultation with and Participation by Affected Populations in Humanitarian Action, as well as the findings of the *Annual Reviews*, suggests that there is a dearth of good practice in involving the affected population in the delivery of humanitarian aid (ALNAP 2003: Draft Practitioner’s Handbook and Global Study Monographs). Similarly, findings from the three HAP field trials in Sierra Leone, Afghanistan and Cambodia describe the challenge of finding and institutionalising adequate methods of effectively listening and responding to the needs and concerns of affected populations during the course of humanitarian operations (HAP, 2003).

Moreover, based on the content of recent reports from the ALNAP Evaluative Reports Database, it would appear that a significant proportion of field workers in emergencies are still not aware of the existence of the Red Cross/Crescent NGO Code of Conduct – which contains a commitment to downward accountability – even though it has been in existence for nearly 10 years (see RRN, 1994). The synthesis of evaluations in Chapter 3 reinforces this conclusion, noting: ‘In terms of consultation with and participation of primary stakeholders, this year’s reports echo the disturbing story of the past two years: the limited ability of agencies to promote participation beyond implementation activities.’ It therefore appears that good ideas and intentions about accountability to affected populations have yet to result in widespread good practice. This in turn is mirrored by poor evaluation practice, as evidenced in the cumulative three year findings of the *Annual Reviews* that 86 per cent of reports were rated as unsatisfactory in terms of consultation with, and participation of, primary stakeholders in the evaluation process. As the meta-

evaluation section notes ‘Despite some good practice, EHA could rightfully be accused of systematically ignoring the views and perspectives of primary stakeholders ... This undermines its credibility and continues in the vein of treating primary stakeholders as passive recipients of aid rather than active participants in their own recovery.’

A fundamental issue in terms of changing mind-sets within ‘traditional’ humanitarian action relates to the extent to which those involved in monitoring, strategic review and RTE are willing to enter into dialogue with primary stakeholders. Usually rushed evaluation exercises allow limited time for such dialogue, while ongoing monitoring over a period of several months offers the opportunity for longer term contact, with repeat visits to the same site, household or individual which should lead to both improved situation analysis and understanding of impact.

A key question is thus whether monitoring systems can be designed to support downward accountability given agency structures that are usually hierarchical in addition to the current extractive function of monitoring with information normally flowing from the field to country offices and thence to HQ. This is not necessarily about replacing specific monitoring activities but rather about being clearer about what monitoring is intended to achieve and how, and changing mind-sets about what monitoring can and should accomplish. Currently, monitoring activities tend to be a repository for a wide assortment of implicit and explicit aims and objectives related to learning and accountability. When problems arise, there is a tendency to call for ‘better monitoring’, without analysing the nature and content of current monitoring and the resources required to shoulder additional tasks, and/or the necessary resources to do what we already do, but well.

Greater downward accountability may mean that monitoring moves further in the direction of RTE and strategic review, as introduced earlier. Of course there are unanswered questions as to how RTE differs from strategic review, and how they both differ from monitoring. This will depend on how an agency defines the scope of monitoring itself. Since monitoring is usually associated with direct management tasks, especially data collection, the difference between these approaches may depend on how real-time evaluators and those involved in strategic review are able to take a step back from day-to-day reporting and administration and use their time for analysis, making sound judgements, and involving primary stakeholders.

## A System under Strain?

1.5

In order to consider the idea of ‘monitoring as a missing link’ it is necessary to look at the wider picture of how monitoring fits into broader humanitarian structures as well as into the task environment of field staff. As this *Annual Review* illustrates, the big picture is one of humanitarian systems under great strain. Two key points can be identified. First, there is an increasing internal reporting burden on field staff and agencies due to multiple reporting demands, increasing earmarking from donors, and a proliferation of cross-cutting themes (e.g. human rights, gender equality, environment), all of which are important but all of which bring their own reporting requirements. It is probably the case that the ‘accountability lobby’ has also added to the burden. Anecdotal evidence suggests that multiple accountability initiatives are viewed with trepidation by field staff – not necessarily because they will reveal malpractice but because they could lead to a time-consuming round of additional workshops and reporting. In the end these initiatives may be more about repackaging field-level knowledge as ‘quality assurance’ information for donors and HQ rather than as useable support for addressing genuine dilemmas of practice.

Second, there are significant problems in enhancing monitoring in relation to human resource practice, especially in relation to excessively long hours of work and high levels of staff turnover, both of which mitigate against providing the time needed for staff to be able to contribute to, and learn from, the information produced by monitoring. Illustrative of the extremes that can be reached is the case of Oxfam in Angola, where over 32 international staff filled a total of 11 posts – an average of three incumbents per post per year. Many other agencies have had similar experiences as detailed in the themed chapter in last year’s *Annual Review*, as well as, for example, a total of 12 evaluative reports this year that note the negative effects of staff turnover on performance (see Chapter 3). In such situations monitoring is less about learning than it is about damage control, as it may be the only way to ensure that incoming staff have some way of understanding what is going on.

The failure of good ideas about accountability and learning to result in widespread good practice may be indicative of blindness to the kind of pressurised work environments that humanitarian aid workers actually experience. It is within this context that genuine incentives to learn – through monitoring and other means – must be put into place. Instead of coming up with additional tasks there is a need to look at how people try to solve problems and make sense of their situation within

prevailing duties and responsibilities. The overwhelming quantity of information and reports that many offices have to produce may reflect a lack of awareness of the actual pressures of fieldwork in terms of poor communication between field and HQ. This theme is also highlighted in Chapter 3

There is also a danger that increased investment in information flows is not sufficiently related to how that information will provide the knowledge needed to deal with the complex and dynamically changing situations that define the working environment of the humanitarian field worker. Operational staff are frequently held accountable based on their capacity to provide more, but not necessarily more useful, information. The concern is that increased information flows may be at the expense of efforts to help transform information into knowledge, especially to embed efforts in an awareness of the context in which affected people struggle to survive and the potential impact of aid on their very survival. It is such wisdom – that is, the ability to apply knowledge to practical action – that the humanitarian sector needs to foster.

In this respect there seems little point in developing yet more methods and/or toolboxes if they are not preceded by a concomitant effort to streamline existing data collection responsibilities. While information needs will vary according to different organisational cultures, contexts in which monitoring is happening and the balance each agency requires between reporting for upward and downward accountability, there are positive signs. Chapter 2 notes an emerging awareness among donors and operational agencies that harmonisation of reporting requirements are needed to free up resources for other tasks. This must be a first step. Learning can only be improved if field staff are given the time to do it. The key message is one of sector reform.

## Linking Relief, Rehabilitation and Development (LRRD)

1.6

Improvements in monitoring for both learning and downward accountability depend to a large degree on the expectations as to what humanitarian assistance sets out to achieve in the first place. Humanitarian programming today, for example, in

Southern Africa, Afghanistan, Sierra Leone, and Ethiopia, is increasingly intermingled with rehabilitation. This implies that humanitarian assistance is no longer only about saving lives but must also relate to the root causes of conflict and poverty. The sector is being called upon to deliver aid in a way that can ultimately reduce violence and promote recovery, development and peace. This ambitious agenda has major consequences for learning, downward accountability and performance. One hesitates to talk about paradigm shifts, an already overused term, but if these aims are genuinely to be realised, monitoring needs to become more than just a vehicle for upward accountability and must be redirected towards providing contextual information to fill information gaps that currently exist between the relief, rehabilitation and development phases of response.

*ALNAP Annual Reviews 2002 and 2003* highlight that a significant proportion of humanitarian funding is actually being spent on activities normally associated with rehabilitation and development and, as Chapter 3 points out, evaluators are increasingly paying attention to this shift. Yet while the sector is beginning to ask the right questions as to LRRD, there are major deficiencies in finding the right answer. The synthesis in Chapter 3 focuses on this theme. It reveals an overall picture of short-term success in most direct emergency interventions, especially in health, water and sanitation, and food aid, but a failure to link short-term objectives with any real lasting benefit. For example, food aid may feed the hungry and save life in the short term but food-for-work schemes appear to provide little lasting benefit. Likewise, clean water is provided in camps but longer term maintenance by community groups and spare parts for pumps are lacking when it comes to providing water on a more sustainable basis. Furthermore, there is little or no sign of capacity building in humanitarian response, including of government and local institutions. As a consequence the most vulnerable are more likely to slip back into destitution when the initial phase of the relief intervention are over.

Perhaps the basic issue is whether it is realistic to expect longer term and more sustainable impacts from emergency interventions? LRRD policies paired with short funding cycles and so-called sunset clauses have created pressure on agencies to make unrealistic claims about the prospects for recovery in order to 'declare victory' and move onto the next humanitarian crisis. The move to results-based management in the sector has supported this tendency towards results inflation. In reality, the most vulnerable are being left behind as the rhetoric of LRRD moves ahead. When struggling to rebuild their livelihoods, the destitute, the disabled and

the landless often lack the resources to keep up with the project cycle. And, at the same time, agencies tend to ignore the coping strategies and capacity of affected populations.

The failure of LRRD on the ground is mirrored by the failure to find a useful synergy between the relief and development communities. Development actors who have knowledge about the nature of ongoing vulnerability and risk are still not engaging with the humanitarian agencies that ‘parachute in’ when, for example, structural food insecurity turns into acute famine. Recent experience in Southern Africa has taught us that chronic vulnerabilities caused by HIV/AIDS and changing political economies mean that the boundaries that are supposed to separate relief and development are becoming fainter still. Understanding the nature of chronic risk is something that needs to be addressed at various levels – namely research, practice and policy.

It may be postulated that this is due to a continued failure, within the humanitarian sector, to look more closely at the wider impacts of the aid provided on the lives of beneficiaries. Downward accountability means more than caring about whether the food was delivered and the bellies filled. It means caring about livelihoods too. Humanitarians may only have relatively blunt tools at their disposal with which to ensure that people are not hungry tomorrow, but that does not absolve the sector from the need to sit with development actors to discuss what needs to be done. And it may be that monitoring has an important role to play in providing information and feedback at key moments in the LRRD process.

## A Summary of Contents

1.7

Chapter 2 begins by providing a brief overview of current monitoring frameworks and practices in the sector, as well as a provisional exploration of the links between monitoring and learning. It then covers the areas discussed in the last three sections of this introduction in relation to potential areas of gap filling, highlighting good practice – improving information flows, simplifying systems, promoting joint activity, focusing on process and impact – and strengthening downward accountability.

Chapter 3 provides the annual synthesis of EHA reports that were made available to the ALNAP Evaluative Reports Database in 2002 (55 reports). The chapter is organised around the main humanitarian action sectors of food aid, water and sanitation, health, and shelter and housing. Supporting sectors are considered in a table. In addition the chapter covers cross-cutting themes, including human resources and management, participation and consultation of affected populations, protection and human rights, gender equality, and results-based management.

The conclusions from this year's *Annual Review* are drawn together in Chapter 4, which stresses the continued importance of the agenda for change and action recommended in *Annual Review 2002*. In addition, some recommendations are made for reviewing and streamlining monitoring systems of humanitarian agencies. The coloured section of the *Annual Review* contains the meta-evaluation, ALNAP's annual assessment of the quality of the previous year's evaluation set. This meta-analysis is achieved through assessment against the ALNAP Quality Proforma (QP) which has been revised this year in light of experience from the 2002 meta-evaluation. In addition to this year's evaluation analysis there is a comparative analysis of the 127 evaluative reports which have been assessed in the meta-evaluation exercise since the *Annual Review* series was launched in 2001, based on comparable questions from the QP used across the three years.

## Annual Review 2004

### 1.8

Next year's *Annual Review* will analyse what happens when humanitarians talk to development actors and primary stakeholders about what needs to be done in humanitarian action. It will focus on field-level learning for improving the understanding of the contexts of humanitarian action. Analysis will draw on ALNAP's experience of testing the LSO concept (Box 2.8), which takes a proactive operational approach to promoting and facilitating opportunities for field-level learning. Additional input will come from the findings of the ALNAP Global Study which has raised important questions about how far agencies are willing to go in operationalising their commitments to learn from affected populations.



## Chapter 2

# Strengthening Monitoring in the Humanitarian Sector<sup>1</sup>

## Overview

### 2.1

What opportunities are there for monitoring to 'fix' some of the ongoing problems in the humanitarian sector? This chapter addresses this question, based on a preliminary research project carried out by ALNAP between September and November 2002. The purpose of this project was to analyse the current state of monitoring and to explore its potential for improving performance, as well as enhancing learning and accountability. As perhaps the first sectorwide assessment of monitoring in the humanitarian sphere, it has allowed general conclusions to be drawn, a review of current innovative practice, and the subsequent development of a set of sectorwide recommendations.

For some time there has been a malaise in elements of the evaluation of humanitarian action (EHA) community concerning the potential for EHA to promote both learning and accountability. As *ALNAP Annual Review 2002* argued, learning functions may be better promoted through internal than external agency functions. And the ability of EHA to hold agencies accountable is at best uneven – not because good evaluations are not being carried out but because mechanisms for fulfilling evaluation recommendations that can overcome political opposition to change in agencies are often not in place. This is not to suggest that EHA does not have its role – mainly 'upward' reporting to executive boards and donors, in addition to promoting emergency-to-emergency learning.

One result of this malaise has been experimentation with assessment processes that complement EHA and support its learning and accountability functions. These new assessment processes have tended to be experimental and uncoordinated – hence the sectorwide review carried out for this chapter. In general the new processes are also characterised by a less formal approach than EHA, and as such monitoring in the humanitarian sector has moved in the same direction as the overall evaluation field, which in recent years has seen a shift to multiple evaluation approaches, and more quantitative evaluation being complemented by approaches which focus to a greater extent on process – for example, utilisation-focused evaluation (Morabito, 2002; Patton, 1997) and empowerment evaluation (Fetterman, 2002).<sup>2</sup>

This chapter is organised as follows. Section 2.1 provides background to the ALNAP research project and sets out some definitions. Next, current monitoring frameworks and practice are analysed to set the context for the remainder of the

chapter, which focuses on ways in which monitoring can be strengthened to meet some of the needs of humanitarian action. Recent innovative approaches to monitoring are explored throughout the chapter, and also used as a basis for discussion of future options.

The consequences of retaining the status quo as far as monitoring is concerned – that interventions are carried out with too little information, and that there is both limited accountability and learning – are too high. This chapter shows that investments in monitoring could have a significant pay-off in terms of improved performance in the humanitarian sector, understanding of processes leading to successful interventions, and downward accountability to primary stakeholders.

## The Research Project Process

### 2.1.1

This chapter is the culmination of the following process:

- An initial literature review and discussion paper, circulated to a peer review group of ALNAP full members, consultants, and other humanitarian practitioners.
- Meetings with operational and specialist monitoring and evaluation staff in 11 ALNAP member agencies, and telephone interviews with other ALNAP members, and consultants.<sup>3</sup>
- Reviews of agency monitoring, reporting and evaluation guidelines, and of relevant evaluations held by the ALNAP Evaluative Reports Database.
- A draft paper presented at the ALNAP Biennial meeting in Delhi in November 2002.
- Discussions in Malawi with ALNAP'S Learning Support Office (LSO), and with NGO staff involved in the joint relief programme in response to the 2002 Southern African drought.

In order to encourage frank discussion of the strengths, weaknesses and potential of current monitoring systems this chapter focuses for the most part on generic rather than agency-specific issues. As agencies still vary widely in their ability and

willingness to be self critical, we wanted to avoid the perverse tendency by which those agencies which are most open in their self-criticism appear to have more problems with respect to monitoring than those which are more cautious.

## Definitions

## 2.1.2

The research project uncovered a monitoring world that is highly fluid, with multiple definitions, approaches and opinions. While there may be substantial sectorwide agreement on the meaning of EHA, monitoring currently has more chameleon-like features. The study found uncertainty about where monitoring fits in agencies' thinking and practice, and in some cases a lack of clarity about the meaning of the term.<sup>4</sup> Moreover, agency guidelines on monitoring and evaluation

### Box 2.1 Definitions of Monitoring and Evaluation

**Evaluation of Humanitarian Action** is defined by ALNAP as:

'A systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability. It has the following characteristics:

- It is commissioned by or in cooperation with the organisation(s) whose performance is being evaluated.
- It is undertaken either by a team of non-employees (external) or by a mixed team of non-employees (external) and employees (internal) from the commissioning organisation and/or the organisation being evaluated.
- It assesses policy and/or practice against recognised criteria: efficiency, effectiveness/timeliness/coordination, impact, connectedness, relevance/appropriateness, coverage, coherence and, as appropriate, protection.
- It articulates findings, draws conclusions and makes recommendations.'

A working definition of **monitoring** is:

- 'The systematic and continuous assessment of the progress of a piece of work over time ... It is a basic and universal management tool for identifying the strengths and weaknesses in a programme. Its purpose is to help all the people involved make appropriate and timely decisions that will improve the quality of the work.' (Gosling & Edwards, 1995:81)

tend to say more about the latter than the former. Alongside monitoring and evaluation are a host of other approaches, including strategic review, several of which are highlighted in this chapter.

General conclusions from the research about agency definitions are broadly summed up in the words of an OFDA staff member: 'Monitoring can be viewed as a program review that gauges progress towards objectives (as well as the relevance/appropriateness of objectives as time/context evolve) and is conducted during the life of the program. On the other hand, we tend to see evaluation more as a post-program impact analysis.' To this can be added the idea that monitoring feeds back information on an ongoing basis to intervention participants in order to promote learning. The ALNAP definition of EHA and a working definition of monitoring are provided in Box 2.1. The key difference between monitoring and evaluation is that the former is usually an ongoing, internal function aimed at providing concurrent knowledge, while the latter is usually a one-shot external function that reports after the event.<sup>5</sup> The core of monitoring is the capacity to collect and manage information. In practice of course there is often overlap and complementarity between monitoring and evaluation. Also, the definition of monitoring in Box 2.1 should be considered in the context of the fluid situation in assessment of humanitarian action as the sector moves to experiment with new ways of learning.

## Current Monitoring Frameworks

### 2.2

Research found a number of monitoring systems in operation across the humanitarian sector, with the type of monitoring carried out dependent to a large extent on agency roles and approach, whether monitoring planned to assess context and/or performance, and the type and stage of emergency.

Agencies have different roles which determine what kind of monitoring they carry out. The needs of agency staff in terms of training and capacity building varies according to the scope of their responsibilities. Monitoring usually focuses on either assessment of performance, or situation analysis. The existence of multiple roles is perhaps one of the reasons why there are a large number of monitoring approaches which aim at assessing performance, as summarised in Table 2.1.

Most UN agencies and NGOs work through a wide variety of implementing partners. In relation to monitoring they tend to be both donors and fund recipients, and both contractors and sub-contractees. Many agencies therefore have multiple roles in that they are likely to be simultaneously monitoring their own work and those of their implementing partners, at the same time as being monitored by their

**Table 2.1** Current Performance Monitoring Approaches

| Activity                        | Definition  | Example   | Methods   | Current Accountability Emphasis |
|---------------------------------|---|---|---|---------------------------------|
| <b>Institutional monitoring</b> | Assessment of management, communications, and human resource functions  | Communication between HQ and field leads to a coordinated response                    | Systems review  | Upward                          |
| <b>Input monitoring</b>         | To check that resources (human, financial, material) are mobilised as planned   | Bags of wheat are loaded, unloaded, and stored  | Mainly quantitative   | Upward                          |
| <b>Output monitoring</b>        | To check that services are being delivered as planned   | Bags of wheat are delivered to primary stakeholders                                   | Mainly quantitative   | Upward and downward             |
| <b>Process monitoring</b>       | Reviews processes by which change takes place as a result of an intervention  | Survey of primary stakeholder perceptions of intervention performance                 | Usually qualitative – e.g. PRA  | Downward                        |
| <b>Impact monitoring</b>        | To verify that the intervention is having the anticipated impact  | Distribution of wheat supports longer term goals such as promotion of gender equality | Quantitative and qualitative, e.g. anthropological surveys, large scale household surveys | Upward and downward             |
| <b>Self-evaluation</b>          | Examines the implementation of an operation in terms of the effectiveness and efficiency of its inputs, activities and outputs. | Country-level programming   | Quantitative and qualitative  | Upward and downward             |

donors. Equally they may be both commissioning reviews and evaluations and being evaluated by their donors.

Table 2.1 is not intended to be comprehensive, and there are other monitoring approaches not included that go by different names. Indeed, terminology often

| Table 2.1              |  | Continued   |   |                                 |
|------------------------|--|---|---|---------------------------------|
| Activity               | Definition   | Example   | Methods   | Current Accountability Emphasis |
| Performance monitoring | Measurement of progress in achieving specific objectives in relation to an implementation plan                                 | Opinion survey as to agency performance   | Combination of questionnaire, qualitative and quantitative  | Upward and downward             |
| Financial monitoring   | Determines whether funds are being used efficiently and as planned   | Audit of agency accounts  | Financial   | Upward                          |
| Strategic review       | Assessment by insiders at one point in time of the progress of an intervention. Takes place during the course of the operation | Investigation of one key issue of importance  | Varied depending on circumstances   | Mainly upward                   |
| Real-time evaluation   | Assessment of operations by external person(s)/ evaluators as they are being carried out                                       | Assessment of relief operations in Afghanistan  | Varied depending on circumstances   | Upward and downward             |
| After action review    | Lesson learning within a team directly after operation is over   | Discussions take place at HQ level or in the field directly after an emergency intervention | Team discussion using three questions: what went well, what did not go so well, and what could be done better next time | Mainly upward                   |

crosses over and some agencies use the term evaluation for a particular set of activities where others use monitoring. However, three conclusions can be drawn from the table:

- 1 accountability is for the most part upward to meet agency needs, rather than downward to primary stakeholders;
- 2 methods tend to be either quantitative or a mix of quantitative and qualitative;
- 3 of the multiple monitoring approaches, many are overlapping. Along with each approach and each agency comes monitoring guidelines and manuals. This in itself is part of the problem – each individual agency has developed its own system and approach, leading to a lack of harmonisation, over-complexity, and multiple monitoring requirements from different donors. Given the multiple approaches as well as the different responsibilities of agency staff, the picture is one of considerable complexity. As such there seems little purpose in introducing new monitoring toolboxes onto an already creaking ship.

## Situation Monitoring

### 2.2.1

Table 2.1 deals mainly with performance monitoring during an emergency. UNICEF (2002) has defined situation monitoring as follows: ‘Situation monitoring measures change in a condition or a set of conditions or lack of change ... It also includes monitoring of the wider context, such as early warning monitoring, or monitoring of socio-economic trends and the country’s wider policy, economic, or institutional context. In contrast, performance monitoring measures progress in achieving specific objectives and results in relation to an implementation plan.’

The different foci of situation and performance monitoring are set out in Figure 2.1.

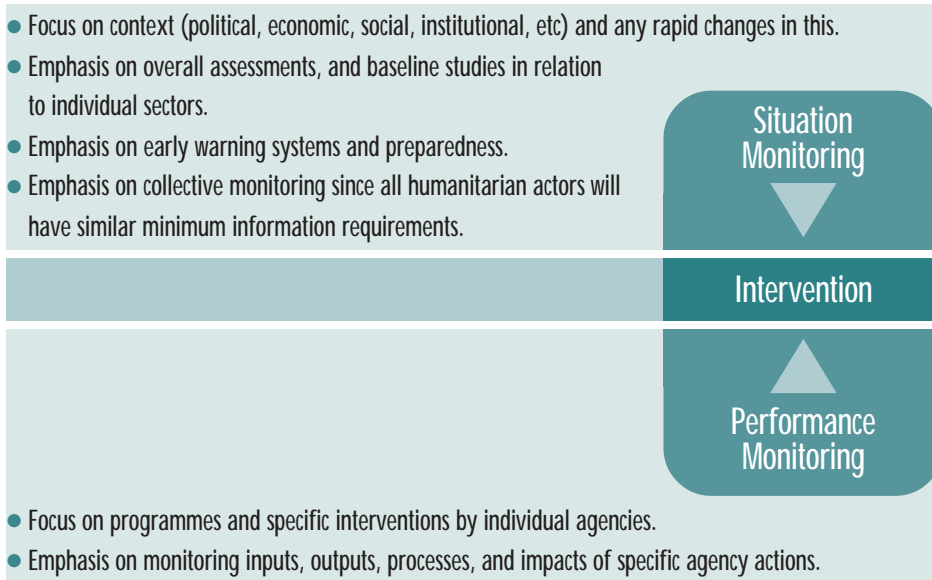
Among the best known forms of situation monitoring is the establishment of early warning systems (EWS), many of which are on-line. Examples include the FAO’s Global Information Early Warning System and USAID’s Famine Early Warning System.



As well as providing impending warning of an emergency, situation monitoring has the important function of providing information that feeds into ongoing monitoring systems during the emergency, including baseline data. Most EHA notes a lack of baseline data against which to measure intervention progress; the following comment from a UNHCR evaluation in Kenya (UNHCR, 2001) is representative of this: ‘While much recognition was given to the importance of baseline data and monitoring, particularly in the Stakeholders’ Seminar and in early project documents, the baseline studies were carried out too late to effect the project design and mode of operation.’

Situation monitoring is also closely tied to needs assessment of primary stakeholders. Much of this needs assessment has focused on combining quantitative and qualitative assessments in an attempt to understand vulnerability. Examples include Save the Children’s (UK) expertise in mapping vulnerability and early warning through the Household Economy Approach, which is intended to enable livelihoods based poverty analysis; also WFP’s Vulnerability Assessment Monitoring, part of its overall emergency needs assessment. This is currently carried out in over 50 countries and uses satellite images of rainfall and crop conditions, and food prices in

**Figure 2.1** Situation and Performance Monitoring



**Table 2.2** Monitoring in Acute Crises and Chronic Emergencies

| Acute Crises   | Chronic Emergencies  |
|--|--|
| <b>Management of M&amp;E</b>   |  |
| <p>Monitoring has to provide very high-frequency and high coverage of information, though not necessarily high precision. Typically in an acute crisis, national information systems deteriorate; previous baselines are invalidated by changes in context. The number of actors increases, making coordination and standardisation of data collection more difficult. Different information sources and data collection methods must be pulled together, compared and analysed, in order to build a national picture – both for situation and programme monitoring.</p> | <p>Characteristics of good management of monitoring will generally be the same as in an unstable context, though planning horizons are often shorter.</p> <p>Monitoring has a key role to play in helping to adjust programmes in a context that remains fluid, such as signalling worsening as well as improving situations; differentiating vulnerabilities; revealing slow effects of crisis; and opportunities for longer term programme strategies.</p> |
| <b>Situation monitoring</b>  |  |
| <ul style="list-style-type: none"> <li>● Rapid assessments.</li> <li>● Reliance on secondary information from partners.</li> <li>● Efforts to include tracking of inaccessible areas.</li> <li>● Establishment of priority plan for future data collection.</li> </ul>   | <ul style="list-style-type: none"> <li>● As in stable contexts, but more frequent updates.</li> <li>● Rapid assessment for localised crisis.</li> </ul>  |
| <b>Programme monitoring</b>  |  |
| <ul style="list-style-type: none"> <li>● Agreement on minimum requirements (key requirements are agreed and prioritised).</li> </ul>   | <ul style="list-style-type: none"> <li>● Monitoring expanded to track priority outputs and outcomes.</li> <li>● Integrated monitoring plans might be re-established.</li> </ul>  |
| <b>Data sharing &amp; coordination</b>   |  |
| <ul style="list-style-type: none"> <li>● Ensure mechanism to coordinate information sharing and collection across sectors within the organisation and among partners.</li> <li>● Minimal guidelines for standardising information with partners are promoted/maintained.</li> <li>● SPHERE standards are used where local indicators are no longer feasible.</li> </ul>  | <ul style="list-style-type: none"> <li>● Same as for stable contexts, with care needed to maintain standardisation of data collection despite turnover of staff.</li> </ul>  |

local markets. This is complemented by participatory surveys to develop indicators of food insecurity and vulnerability.

## Monitoring in Different Types of Emergencies

### 2.2.2

A further important distinction to make in terms of current monitoring frameworks is the kinds of assessment that take place during different stages of the emergency, and between different types of emergency. This is well summarised in Table 2.2, which is adapted from UNICEF (2003). As can be seen in the table, monitoring in acute crises as opposed to chronic emergencies requires different timing and skill sets, with monitoring during chronic emergencies being quite similar to ongoing development monitoring.

2

### Monitoring in Acute Crises

The early stages of an acute crisis is characterised by the very strong demand for information, from both HQ and sometimes the media, and hence a strong degree of (usually informal) feedback to those supplying the information. Equally in acute crises much effort goes into situational assessment and most monitoring is of inputs (e.g., the amount of food or other relief materials delivered) and outputs (e.g., the number of people being assisted).

Most head offices of agencies felt that their monitoring systems in acute crises worked reasonably well, mainly because of strong interaction between HQ and field staff and the widespread use of informal means of monitoring – especially phone calls, e-mails, and visits. It would need more in-depth research to establish whether regional, country, field offices and frontline workers of these same agencies have the same view. The need to ensure that this informal flow of information is documented is dealt with below. The apparent contradiction between agency staff perceptions here, and the understanding that monitoring systems in general do not work well (see below), is perhaps explained by the difference between formal/routine monitoring, which is seen as unproductive, and informal monitoring and communication.

A key issue raised in relation monitoring in acute crises is the degree of trust between field staff on the ground and their colleagues in country, regional and head offices: shared values, good communications, and strong personal relationships are

necessary for the success of any operation, but are especially vital in this stage of the response. This is dealt with in more detail in Section 2.4.1.

### Monitoring in Chronic Emergencies

A key feature of monitoring in chronic or longer term emergencies is the importance of a strong situation analysis as well as the ability to constantly update and refine this. Many long-term emergencies now last longer than development projects, and the major challenge posed by such emergencies for monitoring is sound situation monitoring and well-managed sharing of information between agencies.

A current key feature of acute crises is that there is usually very little time for data collection and analysis, and initially decisions often have to be made on the basis of limited information. In this context it may be unrealistic to expect full baseline surveys to be undertaken before life-saving measures have been put in place. However, in chronic emergencies, like the current drought in Southern Africa, one would expect humanitarian agencies to have excellent baseline data – either through their own contacts in the region, or through their implementing partners on the ground. Yet the evidence from both EHA and staff currently involved in humanitarian work is that they still face difficulties in undertaking baseline assessments, accessing other relevant baseline information, and documenting the results in a form which can be used for later monitoring. Baseline studies also commonly fail to utilise available information from development agencies and academic institutions with long-term experience in the area.

One finding of this research is that the quality of the initial, or baseline, assessment will be influenced by how an agency manages the relationship between humanitarian and development work. Ideally, an agency working in an area vulnerable to frequent floods or periodic droughts will have in place a clear picture both of overall population figures and of the most vulnerable areas and groups of people within that general population. Ideally also, it should be regularly feeding such information about the changing vulnerability of the local population into a national or regional EWS and sharing it with other governmental and non-governmental agencies working in the area

In reality the presence of an agency in a locality in no way guarantees that it will have such data at hand. Overall situation monitoring or early warning mechanisms

tend only to be given priority for about a year after a particular emergency – beyond which time they no longer seem a priority for resource allocation. This partly relates to donor fatigue, as it is difficult to maintain the profile of chronic emergencies when there are competing demands for attention – currently the case with reconstruction in Iraq. NGOs in particular tend to focus on their own primary stakeholders and rarely keep updated information about the wider area. For these reasons, agencies need to build strong lateral linkages with government structures and other local agencies working in the same area in order to develop an information base about an area and its overall vulnerability to different sorts of humanitarian crises. Where government and other local capacity is weak, it needs to be supported – not only in the area of situation analysis but also in terms of monitoring in general. Unfortunately, as noted in Chapter 3, capacity building is one of the weakest areas of agency intervention, and monitoring does not appear to be an exception here.

## Performance and Constraints in Current Monitoring

### 2.3

It is important to acknowledge that agencies have in some cases invested in attempts to strengthen monitoring through the production of guidelines and training of staff. Some innovative examples are highlighted in Section 2.4. However, a common finding of this project and across the three years of the *Annual Review* is that performance monitoring is not meeting its potential for promoting learning or accountability. Findings from the 165 reports covered in the *Annual Review* to date can be summarised as follows (in this case a report on Kosovo, ALNAP 2001:93): ‘Uniform tracking, monitoring and evaluation systems and approaches were not applied in the emergency, nor was adequate capacity available for these functions.’ Some further quotes from reports this year are provided in Box 3.11 in Chapter 3, and point to a uniform picture where even routine monitoring is not being adequately carried out.

One consequence of lack of adequate monitoring systems is that this seriously constrains EHA. A large number of evaluation reports note that it was not possible

to draw firm conclusions on intervention results because of the lack of systematic monitoring. In the absence of such information evaluators may be able to reconstruct what happened or was achieved by looking at visible outputs or by talking to key actors, but with few systematic records of outputs, processes, or outcomes it is far more difficult for them to build up a complete picture. In agencies with weak overall monitoring systems the extent to which the results of a particular humanitarian operation is documented will usually depend on the quality of individual managers, and whether they are still available to inform the evaluation.

Coordination of monitoring functions is not regularised in the sector. The most important collective instrument in the UN system is the Consolidated Appeals Process (CAP). While consolidated appeals involve a joint analysis of context, and refer to cooperation between agencies in the response, they are still seen primarily as a funding mechanism. A recent evaluation of the CAP suggests there are significant problems relating to strategic monitoring: 'The issue of strategic monitoring of the overall humanitarian needs and impact of assistance is likely to increase in prominence over the coming years. The UN, like other aid agencies, lacks the baseline information and monitoring systems to give an overall and holistic picture of needs in a country, and of the impact of the overall assistance programme in meeting those needs. This is a systemic weakness of the humanitarian world as a whole. Its relevance to CAP, however, is that without such information, a strategy that aims to represent and meet the totality of need in a country will never appear fully credible' (Porter, 2002).

This diagnosis of constraints to monitoring is not intended to be needlessly critical of agencies, many of which are already aware of some of these problems and are making efforts to correct them, but to pinpoint areas for action by agencies both individually and as a network. This section highlights two central constraints that need to be overcome for monitoring to achieve its potential.

## Staff Overload

### 2.3.1

Perhaps the most problematic issue is the increasing demands being made on those whose job it is to collect and analyse information. Details from this year's data set on this issue are provided in Chapter 3. All the organisations contacted during this project mentioned increased reporting pressures. UN agencies and some of the larger INGOs have also increased their own internal reporting burdens. A recent

survey of required UNHCR reports found 16 different annual reports, one quarterly statistical report, plus monthly and weekly situation reports. In Malawi, the Joint Emergency Food Aid Programme (JEFAP) is generally seen as a successful national relief distribution programme, but it involved considerable monitoring, for example: weekly/monthly stock and distribution reports; food security monitoring; on-site monitoring; post-distribution primary stakeholder monitoring and non-primary stakeholder monitoring; quarterly reports; distribution registers, ration cards, meeting minutes and distribution plans. In addition to WFP's requirements, implementing partners had their own monitoring systems. The experience of WFP staff was that there were no clear guidelines on the number of staff required to fulfil particular tasks, leading to an overload of duties for field staff or neglect of some necessary duties such as monitoring. Their perception was that there were too many forms; they were unclear about the purpose of each and how the information was to be used. This led to an overhaul of the monitoring system during the programme (see Box 2.8 on the ALNAP LSO for more details).

2

Agencies, in particular the UN agencies, are also being required to monitor against a range of international protocols and commitments, for example gender equality and human rights. One of the major contemporary issues in humanitarian work is the protection of civilians, and frontline humanitarian workers are in the best position to monitor and report on human rights violations. The protection agenda greatly increases the range of issues that humanitarian workers are now expected to monitor. A concern for the protection of civilians also opens new dimensions that need to be monitored even in conventional food relief programmes. In many conflicts, distributing food and other supplies to civilians increases the risk of attack on them. Conventional monitoring of the amounts of food distributed does not measure whether such distributions lead to an increase in the number of attacks on the communities concerned. However, once protection of civilians from attack is accepted as an imperative, monitoring must include issues like the impact of relief distributions on people's security (IASC, 2002; Eguren, 2002).

The introduction of the Sphere *Humanitarian Charter* and *Minimum Standards in Disaster Response* has added a new dimension to monitoring. However, once an agency accepts these standards monitoring may have a further purpose – that of compliance. So far the Sphere Project has promoted awareness of these standards and encouraged debate across agencies as to their relevance as well as how they can best be achieved. The standards themselves also offer a useful benchmark against which the performance of different agencies can be monitored. However, recent

studies by the Sphere Project show that even the 20 agencies piloting Sphere often cite Sphere in their project proposals but have been slow to incorporate the standards in their monitoring work. Only 25 per cent of the offices of the pilot agencies had used Sphere in project monitoring, while twice that number used it in project proposals. Sphere (2002) notes that '[P]ilot agencies are aware that monitoring in general is a particular weakness' and suggests both that more project proposals should include a budget line to cover the costs of monitoring and that 'there should be a paradigm shift from regarding monitoring as a largely extractive exercise aimed at providing information for donors, towards a partnership relationship with donors in which monitoring is a joint activity to test and review progress.'<sup>6</sup>

One paradox found by this research was that monitoring appears in the job descriptions of many agency staff (e.g. as many as 80 per cent in the case of UNHCR), and yet the perception of many of those interviewed, including from larger operational agencies and donors, is that monitoring is a relatively low priority. This may partly relate to donors' perception that there are less visible benefits from monitoring than from evaluation, as the latter results in a clear product (a report) that can often be used as a basis for further funding decisions. At the same time field staff in most agencies feel that they are under constant pressure to produce higher quality monitoring in the form of more sophisticated needs assessments, better targeting, better post-distribution monitoring, enhanced accountability to primary stakeholders, and more frequent impact studies. Yet they argue that they are not given the resources to meet these increasing demands.

This increasing reporting burden has also been caused by the shift in funding patterns away from core funding towards supplementary funding and special grants. This project could not quantify the costs of these changes but the process of decentralisation of many donors has resulted in a proliferation in the number of funding mechanisms, with some donors granting funds to the same agency through both their regional or country offices as well as through their HQ. This increases the overall reporting burden. Reporting requirements tend to increase dramatically following an acute crisis; after the 2000 Mozambique floods the local UNICEF office had to prepare 35 reports in one month. ICRC recently reported in its Annual Report for 2001 that: 'Earmarking has become more widespread in recent years, and contributions have often been accompanied by rigorous timetables for the implementation of projects and stringent specific reporting conditions. Experience has shown that the more restrictive the earmarking policy, the more limited the



ICRC's independence and operational flexibility – much to the detriment of the people that the ICRC is trying to help.' Operational agencies have found that the same donors who are critical about the high level of their overheads then also demand reporting in particular formats on specific earmarked grants.

## Data Focus and Quality

### 2.3.2

The second main constraint to improved monitoring relates to data focus and quality. Three main issues arose during this project, which can be summarised as follows:

2

- monitoring systems tend to focus on collecting data at the input and output levels;
- mainly quantitative data is collected;
- data quality is often poor, and poorly analysed.

### Box 2.2 Input and Output Data

'The familiar adage ["you can lead a horse to water, but you can't make it drink"] illuminates the challenge of committing to outcomes. The desired outcome is that the horse drinks the water. Longer-term outcomes are that the horse stays healthy and works effectively. But because program staff know they can't make a horse drink water, they focus on the things they can control: leading the horse to water, making sure the tank is full, monitoring the quality of the water, and keeping the horse within drinking distance of the water. In short, they focus on the processes of water delivery rather than the outcome of water drunk. Because staff can control processes but cannot guarantee attaining outcomes, government rules and regulations get written specifying exactly how to lead a horse to water. Funding is based on the number of horses led to water. Licences are issued to individuals and programs that meet the qualifications for leading horses to water. Quality awards are made for improving the path to water – and keeping the horse happy along the way. Whether the horse drinks the water gets lost in all this flurry of lead-to-water-ship. Most reporting systems focus on how many horses get led to the water, and how difficult it was to get them there, but never quite get around to finding out whether the horses drank the water and stayed healthy.' (Patton, 1997:157-8)

As Patton suggests, the problem of collecting data beyond the input and output levels is perhaps the central current issue in results-based management. Box 2.3 illustrates why this is problematic.

### **Box 2.3** Knowing What to Monitor

Watkins (2002) assessment of emergency food aid in Kenya found the following: 'A comprehensive information system was developed for monitoring and tracking the large quantity of food commodities in the SFP [Supplementary Feeding Programme]. Accountability of the commodities, from dispatch in Nairobi, to the Baringo warehouse and finally to the distribution site, was accurate and comprehensive. Food monitors, present at each of the 250 distribution sites, reported on the amount of food received and distributed, the time and process of distribution, etc. The information management system allowed the programme to easily quantify losses and damages, and determine the amount of food that was delivered to each family. However, monitoring in the SFP was based almost entirely on commodity tracking as opposed to assessing the use of the food and nutritional status of children etc. There was a risk that the SFP was merely a logistical programme, not very different from the general food distribution itself. While monitoring the nutritional status of individual children may not have been feasible, more regular assessment of the health and nutrition status of the children was required. The large amount of resources that was allocated to food and logistics justifies a much larger component of impact-related monitoring. Monitoring indicators in this context could have included: nutritional status of children and household visits to monitor use of food, preparation practices, and intra-household distribution patterns.'

Similarly, a consultant interviewed for this project found that: 'An interesting problem that I have seen with UN agencies and NGOs alike is that they tend to direct their monitoring at the activity level and neglect performance monitoring at higher levels ... I recently saw a project that consisted of digging fish ponds for food insecure farmers. There was an output statement but it was to dig 'x' number of fish ponds rather than to provide fish (as protein or as a resource to trade). This turned out to be critical because the target number of fish ponds was achieved but none were ever stocked with fish. The monitoring reports thus gave no indication that nothing had been achieved in terms of fish resources or food security (the unstated output and purpose respectively).'

### Bias towards quantitative data

The quote from Watkins (2002) in Box 2.3 is illustrative of a further problem in monitoring: the tendency to ignore the importance of qualitative data. A particular problem identified by this project is that while it is relatively easy to collect quantitative data and send it 'up the line', it is far more difficult both to define what qualitative data staff should collect on a regular basis, and to analyse such data when collected. Staff rarely have the appropriate training and skills to use qualitative data in an effective way.

Yet collection and interpretation of qualitative data on an ongoing basis can be a key means of understanding causality and social process, something that is currently largely missing in EHA (see ALNAP 2002:184). An example from this year's evaluation reports (Chapter 3) is the limited understanding of who benefits from food distribution. Evaluation reports reviewed for the *Annual Reviews* point out that agencies need to know more about why interventions succeeded or failed, and who benefited and lost out. The current focus on quantitative monitoring is excluding a deeper understanding and hence decreasing opportunities for learning.

“By providing masses of quantitative data in reports you could distract people from the key issue of whether funding for the reconstruction of housing was the right approach in the first place.”  
Field worker in Bosnia, 1999

### Data quality

A characteristic of current humanitarian work is the growing gap between the supply of, and demand for, high quality information. Do many field workers believe that collecting and collating figures is relatively 'safe', whereas providing qualitative feedback to their supervisors is a more high-risk occupation? Mebrahtu's (2002) study of monitoring and evaluation by INGOs in Ethiopia suggests a disturbing level of optimistic bias in reporting at the project level. While spot checks by senior staff or donor agency representatives may be useful they are no substitute for robust management systems that support high quality data collection. Findings on results-based systems from this year's synthesis of evaluation reports strongly suggest that in some cases too much, and inappropriate, data is being collected.

In the 2002–3 food distributions in Malawi, NGO staff implementing the JEFAP on behalf of WFP expressed a need for more training on interview techniques and

communication skills as they found it difficult to collect reliable data through interviews. The problems they mentioned included (LSO, 2002): lack of privacy and too many crowds; boredom of interviewees if asked too many questions; and a fear amongst respondents that they might lose their food entitlement if they gave a 'wrong' answer. In the light of these pressures one group of staff estimated that only 50 per cent of interviews produced reliable information. Yet very few emergency staff are likely to get much training in interviewing techniques and, as the Malawi case shows, emergency relief distributions may not offer a conducive environment for successful data collection. Language differences and the use of interpreters can make this kind of communication even more hazardous.

## Strengthening Monitoring

2.4

In response to the constraints and weaknesses noted, this section highlights five areas that offer potential for improving monitoring and feeding into more effective humanitarian action. Together they constitute an agenda for a radical reform of monitoring practice.

### Information Flows: improving trust and feedback

2.4.1

Information flows to and from the field were found to be problematic. Monitoring activities currently tend to move information in one direction, with reporting often the driving force leading to data extracted from the field moving to HQ. This extraction process is characterized by partial reporting in the sense that many reports do not reach the country and regional office, or HQ, and the information that does get through may be selective. The ALNAP LSO in Malawi found that WFP implementing partners had problems processing and forwarding reports; reports completed at field level were not received either by the agencies' HQs or by WFP. The upward flow of information may also result in random 'dispersal' (e.g. incomplete datasets, data processed haphazardly, etc). Data can also be concealed – either willingly, for fear of blame, or because, especially in informal monitoring systems, the importance of a piece of information may not be evident to the information holder.

Despite the extensive reporting required, a commonly noted problem in many agencies is lack of feedback from regional or head offices to situation reports, itself a likely disincentive to maintaining a consistent quality of reporting from the field. Monitoring however should not be simply about extracting information but, above all, about circulating information swiftly to those who need it and can act upon it. Much of what has been written about the characteristics of 'learning organizations' also applies to monitoring: monitoring will only meet its full potential in self-critical organisations where the culture of blame is not predominant and where there are effective communication flows (ALNAP, 2002). There are at least two actions that can be taken to improve trust and feedback:

- **Promoting understanding of the multiple uses of data** A key indicator of the strength of a monitoring system is the quality of feedback that those generating the information receive. Feedback has multiple functions. It can be limited to basic quality control and queries about the data submitted, but it can also provide field staff with real incentives to improve the quality of information they are supplying. The more complex are humanitarian operations, and the greater the number of actors, the more important it is to have strong feedback systems throughout the operation. However, it is common for memos that accompany agency monitoring forms to say little about how data will be used. Agencies need to do a better job in explaining use and ownership: staff are much more likely to spend time and effort collecting and reporting data if they understand the value of the work they do, and are confident that data will be used and valued (Simister, 2000). One recent innovative means of improving communication is RTE, which is highlighted in Box 2.4.
- **Promoting a culture of openness** Staff members are often reluctant to write down findings that their supervisors may consider negative, and reports may be affected by a level of optimistic bias (Mebratu, 2002). Much depends on how senior management responds to the reporting of errors; a culture of blame still appears predominant in many parts of the sector.

### Capturing informal monitoring and communication

A key question that arose in the research is the extent to which managers make decisions using data generated by formal monitoring. We define formal methods as mainly written and routine, and informal ones as verbal and opportunistic in the choice of method used. Anecdotal evidence suggests that formal monitoring systems

**Box 2.4 Innovations in Monitoring: Real-time Evaluation**

Pioneered by UNHCR, partly in response to criticism of the agency's performance in Kosovo, real-time evaluation (RTE) is also being employed by WFP and considered by several other agencies, including UNICEF. Groupe URD is also using a similar approach – 'iterative evaluations with mini-workshop' – although as opposed to other examples the assessment process here is external. Real-time approaches resemble a form of dynamic monitoring that are intended to provide swift feedback to ongoing emergency interventions.

UNHCR (2002:1) defines RTEs as: 'a timely, rapid and interactive peer review of a fast evolving humanitarian operation (usually an emergency) undertaken at an early phase. Its broad objective is to gauge the effectiveness and impact of a given UNHCR response, and to ensure that its findings are used as an immediate catalyst for organizational and operational change.' According to UNHCR, the broad objective of RTE is to (ibid): 'gauge the effectiveness and impact of a given UNHCR response and to ensure that its findings are used as an immediate catalyst for organisational and operational change.' Characteristics of RTE are: timeliness (undertaken in the early phase of an operation, at a time when key operational and policy decisions are being taken); interactivity (evaluators are directly involved in the emergency planning process); and perspective (the evaluators are a repository of knowledge on lessons from past emergency evaluations and should incorporate such knowledge into the evaluation process and outputs). In addition, a key element is ensuring quick

either fail to provide information in a form that can be used by managers, or that this information cannot be analysed and summarised in a user-friendly format within the tight timescale required for decision taking. This suggests that formal monitoring data may be more useful for donor reports, and that managers themselves often make their decisions on the basis of more informal consultations with staff or implementing partners.<sup>7</sup>

This is perhaps because formal monitoring systems alone may not guarantee the degree of responsiveness required in an emergency, as they simply cannot cover all the eventualities that can occur. Informal monitoring systems may help identify unexpected problems, opportunities and impacts. As Simister notes (2000:9): 'Field

**Box 2.4 Continued**

flow of information both to the field and to senior managers. RTE is normally conducted jointly by staff from the evaluation department of an agency, or by consultants who have a close relationship with, and understanding of, the agency's operations.

RTE breaks down the boundaries between monitoring and evaluation, and is spurred by the consideration that conventional evaluation is ineffective at providing immediate feedback in emergency situations. There is clearly an overlap between strategic review and RTE; the main difference between them is that while a strategic review can be run by managers themselves, an RTE is usually run by staff from the agency's evaluation unit and possibly outside consultants. The main difference with monitoring is that RTEs are not mainstreamed in programmes. There is a thus risk here of alienating and disempowering staff, and much of the success of a RTE will depend on the sensitivity of the evaluators. Other limitations noticed (ibid) are the short time available for consultation with beneficiaries; the narrow focus upon UNHCR's own operations/absence of a sectorwide perspective; and the risk of placing controversial findings in the open.

Despite these potential constraints, RTE offers an opportunity to both revitalise monitoring and overcome some of the problems with conventional evaluation. Much will depend on the quality of the evaluation department and their ability to package information for both the field and senior levels of the agency.

staff can play an especially important role in identifying this kind of information, particularly where they are in close contact with the intended beneficiaries ... However ... channels of communication need to be established through which information can be passed to higher level managers on an irregular basis.'

By definition, much informal monitoring remains undocumented. But this does not mean that it is unimportant. In most organisations both staff and primary stakeholders are continuously engaged in different monitoring activities – in relation both to the overall situation and a wide range of performance issues. The following response from CAFOD to a question about the use of informal monitoring methods is representative of the sector as a whole: 'Particularly in major

emergencies, CAFOD's emergency staff regularly visit partners to review the progress of their programmes, to discuss any problems that may have been encountered, and to modify plans to ensure that the programmes are successful' (Personal Communication: Matthew Carter 2002).

Because formal monitoring systems are not meeting information needs during emergencies does not mean that adequate information flows are not taking place. However, a strong reporting system needs to find ways of documenting information that would otherwise go unrecorded so that lessons learned can be analysed and disseminated. Events like well-facilitated, and well-documented strategic review meetings and after action reviews can encourage dialogue within an agency and help bridge the gap between the formal and the informal. Although frontline staff may be reluctant to share findings they fear may be interpreted negatively by their immediate supervisors, they may find it easier to share their knowledge of a particular relief operation in a well-facilitated interagency meeting, such as those organised by ALNAP's LSO in Malawi.

## Simplifying Systems

### 2.4.2

A key element in simplifying systems and cutting down staff overload is prioritising information needs. As Borton & Macrae (1997:55) note: 'A balance clearly needs to be struck in terms of the reporting demands placed on relief staff working in often difficult situations, and ensuring that information flows are of sufficient quality to inform future decision-making and to facilitate institutional learning. Central to this balancing act will be donors identifying their priority information needs'. Section 2.3.1 above noted the situation where agencies are required to report on a wide range of international and national commitments; in effect reporting on these commitments is a competing rather than a complementary exercise. Agencies need to decide whether the current situation – reporting poorly on a wide number of commitments – is good enough. If not, a realistic prioritisation into main and secondary areas of reporting should be made by policy units.

Agency departments responsible for monitoring may be located in technical divisions or departments, with a mandate to promote the collection of information to meet HQ requirements. One example is financial monitoring; a number of UN agencies have been developing coding systems to capture allocations to different priorities in relation to international protocol, and UNICEF, UNFPA and WFP, for



example, are revising these systems to make them easier to use, after feedback from the field about staff overload and the complexity of systems. The key in developing systems is often to aim for the bare minimum of necessary information (IANWGE, 2003).

On the side of implementing agencies, UNICEF is proposing that those donors who wish to cherry pick small projects should pay a much higher percentage overhead charge than those supplying less restricted funding. ICRC has negotiated a standard reporting system with its key donors and now explicitly asks its country delegates to review outcomes against a set of expectations documented in an annual plan. USAID/OFDA tries to reduce the reporting burden of its implementing agencies by centralising grant making in Washington, and aims to consolidate different grants to implementing agencies in a single funding 'envelope'. Box 2.5 provides another example of USAID's attempts to simplify systems. The UN Development Assistance Framework and harmonisation process in the UN may also eventually lead to simpler systems and more coordinated reporting.

### **Box 2.5** Simplifying Systems through the Introduction of Internal Opinion Surveys

For donor agencies, internal opinion surveys may be a useful, low cost way of undertaking performance monitoring. This may be especially useful where the recipient organisation works in many different countries. USAID/OFDA has started to seek opinions from its staff around the world on the performance of international organisations, both UN agencies and INGOs, which receive USAID funding. It asks its staff to comment on the outcomes of activities supported in their particular countries using a simple percentage scale, with 100 per cent being a totally favourable impression and 0 per cent a totally unfavourable one. Staff are encouraged (but not required) to provide comments to support the scores they give. The results of such surveys are necessarily subjective but do at least provide an overview of an agency's performance in different contexts. The advantage of such surveys is that they encourage agency staff to make qualitative judgements which are usually lacking in normal monitoring systems, and if repeated, perhaps on an annual basis could provide evidence about whether a group of staff feel a particular agency's performance is improving or declining. (Personal communication with USAID/OFDA staff, October 2002)

## Promoting Joint Activity

### 2.4.3

As part of this project collective monitoring initiatives within the humanitarian sector were reviewed to examine how far these offer the potential for a more effective and streamlined monitoring system. Different perspectives were found, first on what constitutes collective monitoring, and second, whether such collective monitoring is desirable or feasible. Collective monitoring can be defined as: 'a joint assessment of a humanitarian situation or monitoring of a particular operation by a number of different cooperating agencies, sometimes including governments.' This implies collective agreement about objectives, indicators, and methods used for data collection.

All humanitarian agencies retain distinctive mandates, different funding lines, ways of operating, and organisational cultures. And while there may be pressure for more collective action there is equally strong opposing pressure due to the competition

#### **Box 2.6** Humanitarian Information Centres

The introduction of Humanitarian Information Centres (HICs) by OCHA supports the case concerning the potential for promoting a more organised and collective situation analysis across the humanitarian sector. HICs aim to manage information 'to enable individual organisations to improve their delivery of assistance... (and) enable them to share their information with other organisations on a more systematic basis. This in turn leads to accurate information reaching decision makers more quickly, improving the appropriateness and timeliness of the response, and creates a shared frame of reference that enables those decision-makers to co-ordinate their activities, based on clear knowledge of needs in the field and clear understanding of each organisation's capacities' (OCHA, 2002). HICs or their equivalent operate in Iraq, Afghanistan, Sierra Leone, the Horn of Africa, Southern Africa and the occupied Palestine territory. One lesson from the current Southern African drought is the need for HICs to be set up at country rather than regional level if they are to achieve their full potential.<sup>8</sup>

Though there is a hint here of collective monitoring in the idea of a shared frame of reference, HICs are primarily seen as providing an information service both for individual agencies and for the humanitarian sector as a whole. Ideally HICs and other similar resource centres should enable agencies to have far better

for funding and profile between agencies. Even so there are plenty of examples of cooperation, especially in relation to joint assessments. Future research projects should study in more detail the costs and benefits of sustained collective monitoring throughout a jointly implemented programme, for example Operation Lifeline Sudan or the Somalia Aid Coordination Body.

Situation monitoring perhaps offers the best opportunity for joint activity and information sharing to complement individual agency practice, as political issues are less prominent. Work on joint situation monitoring has recently gained momentum with the development of common practice (e.g., the Share Project; King & Dilley, 2001) and common standards and guidelines for information exchange (see OCHA, 2002). However, the 'lack of importance that is given to information management at HQ level in most (if not all) humanitarian organisations' (Currion, 2001) is still a major constraint to the development of shared information initiatives.

#### Box 2.6 Continued

contextual information – at least with respect to geographical and physical infrastructure if not on social, political, and institutional issues.

HICs are now strongly backed by some of the major donor agencies – especially USAID and DFID. USAID makes information sharing a condition of its funding for emergencies, and its guidelines state that 'OFDA will not fund organizations that do not share programmatic data and information with appropriate humanitarian information co-ordination bodies in the field, such as Humanitarian Information Centers' (USAID, 2002). However, discussions with agencies suggests a clear distinction between supplying information services to agencies (which will always be popular) and using this information as a basis for suggesting agencies work in particular areas or sectors (which remains controversial). HICs do offer the potential basis for a collective monitoring system from which agencies would work cooperatively to ensure more consistent coverage of relief and rehabilitation assistance in all areas and sectors. The key questions for the humanitarian sector as a whole concern what level of collective monitoring, even in relation to more cooperation and information sharing on assessments and baseline information, the sector feels is desirable, and what degree of coordination will individual agencies be prepared to accept in practice?

As noted, the most important collective instrument in the UN system is the CAP. The recent review of the CAP makes some pertinent recommendations in relation to monitoring (Porter, 2002): '[D]onors and agencies have a common interest in improving monitoring systems to guide needs assessments and to measure the impact of humanitarian programmes, and should work together to do so.' It also notes that the UN tends to present information about humanitarian need in CAP countries: 'exclusively in terms of unmet funding requirements. While such figures may be useful indications, they must be backed up with accurate field-based indicators of humanitarian need gathered by improved monitoring systems' (ibid).

Promising examples of information sharing and exchange in relation to situation monitoring are presented in Boxes 2.6 to 2.8. The common feature of these examples is that they are organised by a respected coordinating agency.

#### **Box 2.7** Information Management System for Mine Action (IMSMA)

IMSMA is an information management system aiming to improve assessment, monitoring and planning at the country level. The IMSMA software, released in 1999, is the *de facto* standard for the exchange of mine action information and is in use in 28 countries (Geneva International Centre for Humanitarian Demining, 2003). At country level the IMSMA is centrally maintained by the Mine Action Centres (MACs), the national authorities for mine action, which include in their mandate the formulation of national strategies and plans, the collection of mine data and provision of information, and the accreditation of organisations working in the sector. All information deriving from national surveys is fed into the IMSMA, which is then constantly updated with information on ongoing programmes by implementing organisations. The IMSMA can therefore track the progress of mine action activities and the corresponding reduction of the impact of landmines, and can be used to coordinate and prioritise mine action activities. In Kosovo, for example, IMSMA was used to facilitate the coordination of diverse activities, including mine clearance and mine risk education; the contributions of the Kosovo Mine Action Coordination Centre are highlighted in Box 3.10 in Chapter 3. It is important to note here that part of the success of the IMSMA is due to the highly centralised structure of mine action, which is coordinated through the MACs to an extent unparalleled in other sectors (on IMSMA see Benini *et al*, 2001).

**Box 2.8 The Learning Support Office (LSO) Test in Malawi**

The LSO concept is being developed by ALNAP members as a means of improving learning during and from humanitarian operations in order to improve agency performance in that and subsequent operations.

The concept has evolved since its conception during the 1999 Kosovo operations and has been refined through a series of retrospective assessments and market testing studies in the varied operational contexts of Orissa, East Timor and Sierra Leone. The food security emergency operations in Malawi during 2002-03 provided the context for the first operational test of the LSO and a team of internationally and locally recruited personnel worked in Malawi from September 2002 to the end of March 2003. The test project was funded by five ALNAP member organisations and overseen by an Interest Group and a smaller Steering Group of ALNAP members.

The LSO Test involved the promotion and facilitation of three different types of learning activities for the organisations and their staff in Malawi:

**1 'Learning-in'** from previous operations and experience, by means of:

- a well-equipped Resource Centre that delivered requested documents to relief workers;
- the preparation of literature reviews and analysis on critical issues such as the relationship between HIV/AIDS and Food Security; and
- the inputting of knowledge and advice through participation in agency meetings.

**2 'Lateral learning'** between organisations, teams and individuals working in the operation, by means of:

- workshops;
- seminars;
- participation in meetings.
- A particularly fruitful intervention was support to the agencies distributing general rations through a series of workshops for field officers which then used the material generated to prepare a Malawi-specific manual on general food distributions which was then used as the basis for the training of all field officers.

**Box 2.8** Continued

**3 ‘Learning-out’** involving capturing learning from the operation for use in subsequent operations, by means of:

- archiving of documents;
- the facilitation of the work of visiting evaluators and researchers; and
- the sharing of lessons.

Whilst the lessons from the Malawi Test and the future development of the concept were under consideration by the Interest Group at the time of writing, key provisional findings so far identified include:

- an independent, objective and respected ‘learning support’ capacity can benefit ongoing operations;
- ‘lateral learning’ appears to be the most fruitful and well-received type of learning support;
- ‘learning-in’ would be greatly enhanced by the LSO being present during the planning and programme design phase of an operation;
- agencies and their personnel sometimes react defensively towards any external criticism; achieving effective collective ‘learning-out’ will require changes in organisational cultures and/or the use of uncritical approaches such as Appreciative Inquiry.

### Including a Focus on Process and Impact

### 2.4.4

One of the major current gaps in EHA is lack of information on impact and process (ALNAP, 2002). Because it is ongoing through and after an emergency, monitoring offers the potential for capturing information which a brief, one-shot, ex-post evaluation cannot do. The key element here is the existence of staff who are attuned, and have the capacity, to analyse issues related to social process, such as intra-household food distribution and gender relations.

Monitoring may not be able to examine impact and process in all emergencies. In sudden impact disasters with short timescales, for example, monitoring may never move far beyond measuring input and output results. However, the longer an operation continues the greater the importance of moving to monitoring processes and impacts, and of maintaining a realistic analysis of the causes of the problems a humanitarian intervention is meant to tackle. While the evidence suggests that it remains a challenge for agencies to move their monitoring from administrative or logistical issues to those related to process and impact, the return to a shift in focus in terms of designing more effective programmes could be enormous.

In addition, monitoring which focuses on social process may support resolution of one of the thorniest problems of humanitarian action – i.e., the gap between relief, rehabilitation and development (LRRD). The synthesis chapter this year provides an analysis of the serious consequences when the LRRD gap is not bridged in terms of resources wasted. Ongoing monitoring could provide data – or ‘information bridges’ – between the different phases of emergency response that are generally missing because of high staff turnover and/or allocation of responsibility for different phases to different agency departments.

Eguren (2002) makes a similar argument for a shift in focus in a paper on operational frameworks for protection:

Current monitoring and evaluation criteria are not geared to the specific needs of field protection work. Conflict situations are more complex and volatile than is generally recognised, with unstable divisions of power, constrained political institutions, and a civil society directly affected by the conflict with its basic needs unmet. In this context we cannot pretend that evaluation criteria and practice focus only on administrative performance and quantitative results. We must also try to measure the political and social impact which field protection has.

One useful tool for examining process is the International Rescue Committee’s (2001) *Causal Pathway Framework*, which emphasises the importance of the process of intervention planning and the need to think logically about the causal links (or pathway) between inputs, activities, outputs, effects, and desired impact. One value of this framework is that it requires the user to state clearly the hypothesis on which a proposed project rests, and to test this hypothesis with reference to objective data.

One part of the move to a focus on process will be an attempt to balance more effectively quantitative and qualitative data collection and use. While collection of quantitative data – the current focus of monitoring – tells us *what* happened, qualitative data tells us *why* this happened; as such, the two data collection approaches are complementary. However, given that staff are already currently overloaded, any move to introduce requirements for collection of more data should be carefully considered. But agencies should also recognise that unless they understand why interventions succeed or fail they will be condemned to repeat past mistakes.

### Consultation with Primary Stakeholders and Downward Accountability 2.4.5

The last of the five potential areas of improvement relates to participation. OECD-DAC's (1999:26) conclusions four years back still hold today: 'Humanitarian assistance is essentially a "top down" process. Humanitarian agencies are often poor at consulting or involving members of the affected population and beneficiaries of their assistance.' The three years of the *Annual Review* support this conclusion with evidence from 165 evaluation reports. Despite increasing discussion of participation in agency guidelines, demonstrated, for example, by Kaiser's (2002) study, there is little evidence that primary stakeholders are increasing their participation in design and planning; and while participation in implementation is more widespread, this may involve taking part in activities already determined by external agents.

Constraints to participation in emergency situations should not be underestimated. The need to consult with primary stakeholders might even be felt as an additional pressure (Groupe URD, 2002). However, much humanitarian action extends well beyond the emergency phase (see Chapter 3 for details), and even in the emergency situation participation may be feasible where agencies or the government have a history of involvement in an area. The HAP report (2002:28) on its field trial in Afghanistan pointed out: 'A number of the NGOs engaged in emergency activities had devised innovative ways to gain the views of project participants, which included specific complaints mechanisms and regular surveys of their program areas. Agencies were quite candid about the constraints they faced in consulting with project participants and those listed included obstruction by powerful members of the community, lack of time, levels of trust between the community and the agency, and the extent to which communities had a sufficient understanding of project goals and agency mandates to be able to make informed comment.'



**Box 2.9 ActionAid and Downward Accountability**

Introduced in 2000, the Accountability, Learning and Planning System (ALPS) is an attempt to reduce the massive amount of work involved in programme reporting, to reduce reliance on written reports, and to learn more from existing programmes. A key component of ALPS is replacing country reports with annual participatory reviews and reflections: 'By reducing the drudgery of written reporting ... ALPS should make space for staff to interact more with partners and poor people. And by introducing processes of review and reflection, it is intended to help poor people, our partners and ourselves, to learn from our experiences and those of others in order to continuously improve the quality of our work' (ActionAid, 2000:6). Since its adoption, ALPS has apparently been internalised in the organisation and resulted in explicit reflection-learning-action cycles, increased downward accountability, a culture of transparency and better understanding of impact. Monitoring now has more of a participatory than a reporting focus (David & Mancini, 2003).

The introduction of ALPS is clearly linked to an advocacy approach to humanitarian action. Part of this has been to increase information for primary stakeholders. For example, after the Orissa Cyclone in 1999, 'ActionAid – with partner agency BGVS and village reconstruction committees – began a food-for-work programme. Daily work charts in public places detailed the number of people working, the jobs they were doing, how much rice was available for distribution and the schedule for cash payments. "People's hearings" acted like vigilance committees. They enabled jobs to be clarified or questioned, and helped curb corruption [...] ActionAid India and BGVS commissioned a social audit to evaluate the food-for-work programme. In the short term, the audit process helped claimants to voice their complaints and seek redress. But participation by women in the social audit meetings also increased significantly their participation in food-for-work, as well as suggesting ideas and changes for future programmes' (Davidson, 2002:18). Similarly, the evaluation of Disasters Emergency Committee funding to NGOs after the Gujarat earthquake in 2001 pointed out that: 'the only agency to make a strong effort on the issue of involvement was ActionAid which instituted a policy of transparency and invited partners to take part in consultative processes' (DEC, 2002:Vol 2:38).

Can monitoring improve information flows between primary stakeholders and agencies, and promote downward accountability and participation in the crucial needs assessment, design and planning stages of interventions? An example of how this has been achieved by ActionAid is given in Box 2.9.

## An Agenda for Improvement: recommendations

2.5

This first sectorwide assessment of monitoring should be seen as an initial step towards rejuvenating monitoring in the humanitarian sector. Monitoring offers the opportunity for EHA to be complemented by a flexible tool that is participatory and focuses on process – i.e., a process that could fill in some of the main gaps in current EHA. Agencies are increasingly realising this and there is a move towards creative thinking in some quarters involving breaking down the boundaries between evaluation and monitoring. This chapter has highlighted examples of innovative attempts to improve information flows, learning and accountability – for example, RTE, the Humanitarian Information Centres, the ALNAP LSO, and ActionAid's Accountability, Learning and Planning System. All of these and other initiatives offer a base from which to learn and build. The challenge is to continue to document and disseminate new practices on a sectorwide basis, and to test their rigour and usefulness in providing information that makes humanitarian action more effective.

Does this make monitoring the new EHA *wunderkind* on the block? Unfortunately, not yet. Given the conclusions in this chapter about staff overload and lack of prioritisation of information needs by donors and others, it is important not to try and load too much onto monitoring. It could, doubtless, and given adequate resources, fill many of the gaps in EHA and many of the information needs of humanitarian response. As these resources are unlikely to suddenly emerge, recommendations on what can be achieved need to be feasible. At the same time interagency communication and analysis of monitoring issues should continue.

One of the areas that this chapter did not have the scope to cover in much detail is linkages to national capacity, both of government and civil society. A characteristic of

some of the new approaches highlighted in this chapter is that they continue to rely mainly on agency staff and resources, and are therefore unlikely to support sustainable in-country capacity. Just like humanitarian action in general, agencies need to start thinking seriously about ways in which the capacity to monitor and evaluate can be handed to national institutions.

The potential for monitoring to expand its role beyond reporting for upward accountability and contribute to learning, downward accountability, and LRRD depends on many factors. We recommend that operational and donor agencies, and interagency initiatives, consider the following, in order of priority.

**Operational agencies** should, over the next six months:

2

- 1 Ensure adequate resources are made available for effective monitoring, using separate budget heads for monitoring to track allocations.
- 2 Ensure that recruitment guidelines and key competencies reflect the need for staff with appropriate skills in monitoring.
- 3 Review staff incentives (professional advancement, informal encouragement, supervision, training and feedback) for maintaining high standards of monitoring, especially in relation to data collection and analysis.

In addition, over six months to a year they should:

- 1 Review their monitoring systems to examine where monitoring fits in the organisation, and whether monitoring promotes increased dialogue, trust and communication flows to and from the field.
- 2 Review their monitoring systems to assess the extent they enable assessment of process and impact, in addition to inputs and outputs.

**Donor agencies** should, over the next six months to a year:

- 1 Undertake an audit of the planning, monitoring, and reporting structures of their implementing partners to determine whether these systems are congruent with reporting on policy objectives, and the extent to which these

systems can be streamlined (with a focus on excluding collection of unnecessary data) and improving the quality of data currently collected. The audit should assess the costs involved in gathering data that is needed to understand impacts, and adjust financial guidelines and incentives accordingly.

- 2 Assess whether monitoring data requested from implementing partners supports internal learning priorities.

**ALNAP and other interagency accountability and learning initiatives** should, over the next six months:

- 1 Explore the possibility of creating a 'community of practice' on monitoring in order to address the issues raised in this year's *Annual Review* and operationalise the recommendations above.
- 2 Initiate a broad discussion among those involved in the 'accountability architecture', with strong participation from field level actors, on the question of whether 'we are part of the solution, or part of the problem' with regard to the current systems overload for monitoring.
- 3 Commit to more pragmatic analysis of the trade-offs and potential synergies between promoting upward and downward accountability.
- 4 Undertake a specific analysis of how learning initiatives, such as HICs and the LSO, could play a more direct role in (a) enhancing capacity for situation monitoring, (b) coordinating collaboration on joint monitoring initiatives, and (c) raising the status and profile of impact monitoring.

## Chapter 3

# Synthesis of Findings of 2001–2002 Evaluation Reports

## Overview

3.1

### Introduction

3.1.1

This chapter provides an overview of performance in the humanitarian sector, synthesising the main findings, conclusions and recommendations of 55 evaluation reports received by the ALNAP Evaluative Reports Database in 2002.

The chapter is organised by the main sectors covered in the evaluation reports: food aid and emergency agriculture; water and sanitation; health; and shelter and housing. Table 3.1 summarises findings from reports not included in these main sectors. The remainder of the chapter covers cross-cutting themes which include: staffing and human resources; partnerships and capacity building; coping strategies; consultation and participation; tied aid; gender equality; coordination; the environment; rights-based approaches; and results-based planning.<sup>1</sup>

As in the two previous ALNAP *Annual Reviews* there is much to celebrate about humanitarian action. Overall the system is working and aid is getting through in all sectors. But Minear (2002:70) poses an important question: 'Can a relief operation be considered a success if, although the patient survives, the local emergency medical team that will treat the patient's next emergency remains marginalized or incapacitated?' In much evaluation of humanitarian action (EHA) the answer is 'yes': the operation *is* considered a success if the patient survives as this is the way most objectives related to humanitarian action are phrased. Evaluated against short-term objectives, humanitarian action is doing well; but evaluated against wider objectives such as sustainability and connectedness, it is almost invariably weak.

The reports reviewed cover many sectors, areas and issues, and this chapter highlights those that were central. A significant focus of the 2002 evaluation set was sustainability/connectedness and capacity building. In this respect, the reports are littered with images of hospitals built but partly used; food-for-work schemes providing little lasting benefit; and hand pumps unused for lack of spare parts and maintenance. This goes beyond the evaluator's penchant for seeking out bad news stories and points to a systemic lack of capacity building and attention to sustainability. As agencies have been pushed towards a focus on sustainability (see Macrae, 2002) they have developed policies on linking relief, rehabilitation and

development (LRRD).<sup>2</sup> However, guidelines and training necessary for implementation of these policies have not been adequately developed or disseminated. This leaves agency staff in a position where they are likely to repeat past mistakes.

Difficulties with LRRD, including the political convenience of having different budget heads, lack of synergy within and between agencies in terms of planning and finance, and lack of capacity building of national institutions, not to speak of conceptual difficulties, has been the subject of extensive review and debate which will not be repeated here (see Minear, 2002; OCHA, April 2002; ECHO, 1999; Macrae, 2001). Rather, this chapter will substantiate many of the findings from the LRRD debate with evidence from current evaluations, and conclude as to how well agencies are currently performing in terms of the various aspects of LRRD – including capacity building, establishing sustainable programmes, consultation and participation, and local versus international procurement.

### The Sample

#### 3.1.2

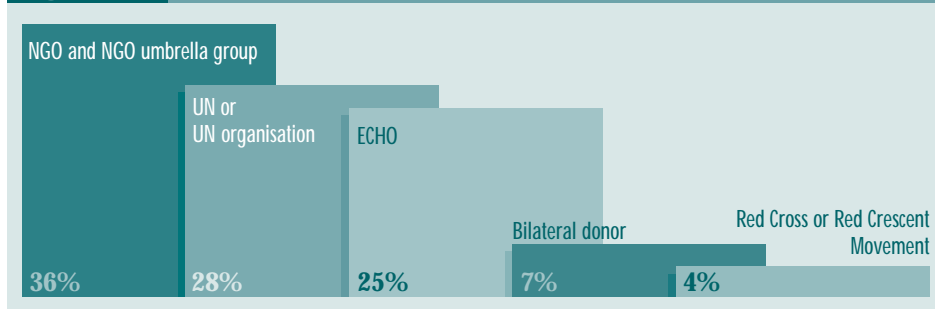
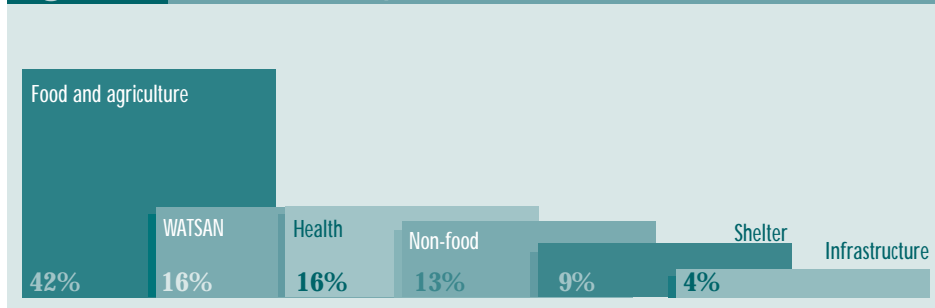
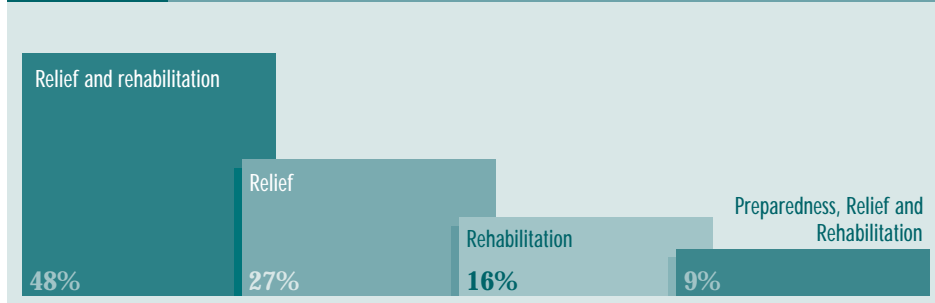
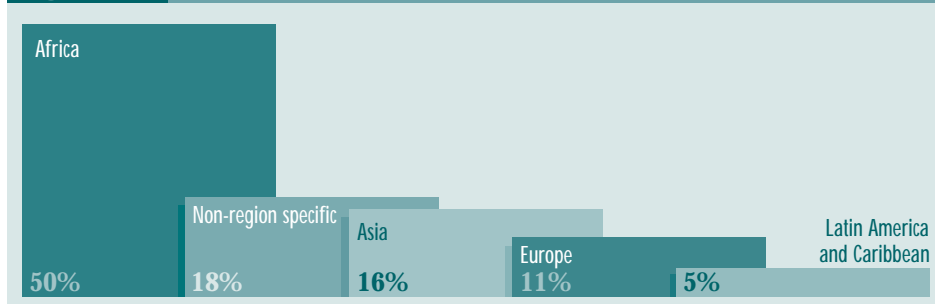
3

Forty-nine individual evaluation reports and six synthesis reports were made available to ALNAP in 2002 (see Annex 3 for summaries of the data set). Fifty-two are independent evaluations, including evaluations carried out by consultants external to the agency and evaluations carried out by mixed teams of internal/external evaluators, and three are internal evaluations.<sup>3</sup> Over three years of the ALNAP *Annual Reviews* this brings the total to 145 individual reports and 20 synthesis reports. This is probably the most detailed assessment of EHA to date, and where possible comparisons over the three years have been made.<sup>4</sup>

### Commissioning Organisations

#### 3.1.3

The breakdown of the evaluation set is shown in Figure 3.1. It is dominated by two commissioning organisations: ECHO and WFP. Ten individual and four synthesis reports were commissioned by ECHO, and eight individual and one thematic report by WFP. The limited number of evaluations provided by the Red Cross/Red Crescent Movement and bilateral donors is apparent.

**Figure 3.1** Commissioning Agencies**Figure 3.2** Distribution of Reports Across the Sectors**Figure 3.3** State of Emergency**Figure 3.4** Regions Covered by the Reports



## Main Sectors

### 3.1.4

Close to 50 per cent of reports focus on food aid and emergency agriculture, with less attention given to housing (see Figure 3.2). A wide range of individual sectors was also covered, including the environment, child protection, mine action and employment generation.

Perhaps half or more of the expenditure was used for relief purposes (Figure 3.3). *Annual Review 2002* noted that between 30 and 50 per cent of funds allocated under the humanitarian budget are used for rehabilitation/recovery purposes. While it is difficult to establish a precise breakdown given the way that expenditure is reported (or in many cases not reported) the impression is that a similar amount was used for rehabilitation/recovery in the reports covered in this overview.

## Countries and Regions Covered

### 3.1.5

A wide and representative range of countries is included. About half the sample (24 of the individual reports and three of the synthesis reports) covers African countries: Angola, Burundi, Democratic Republic of Congo (DRC), Ethiopia, Kenya, Liberia, Sierra Leone, Somalia, Uganda and Sudan. Seven individual reports cover Asian countries: Afghanistan (x3), India, Bangladesh, DPR Korea (DPRK) and Cambodia. Of the remaining individual reports with single country focus, four reports cover the former Republic of Yugoslavia, with one each on Brazil, El Salvador, Azerbaijan, the Caucasus and Iran. The remainder are multi-country in focus. A majority of reports – 32 – cover complex emergencies or countries emerging from complex emergencies; 14 reports cover natural disasters; nine are mixed – i.e., natural disasters occurring in the context of a complex emergency (e.g. the volcanic eruption in the DRC).

## Elaboration of the Main Findings by Sector 3.2

### Food Aid and Emergency Agriculture<sup>5</sup> 3.2.1

This section is based on 18 individual reports and one synthesis report. Of the individual reports, seven were commissioned by WFP as part of its assessment of the Protracted Relief and Recovery Operation (PRRO) mode, covering Azerbaijan, Iran, Ethiopia, Angola, the Great Lakes region, Somalia and Uganda.<sup>6</sup> Four reports were commissioned by ECHO with a specific focus on food security or nutrition: Sierra Leone, DRC, and two reports on Burundi. Six reports commissioned by NGOs included details on food: by Tearfund in Afghanistan and Burundi, the DEC in Gujarat, World Vision in Kenya, CARITAS in DPRK, and Handicap International in Cambodia. The remaining individual report is a joint World Vision/UNICEF evaluation of operations in Afghanistan. The FAO report covers preparedness, emergency agriculture, LRRD and strengthening resilience in disaster prone countries. It is based on a 15-country study, a questionnaire, and project document review.

This large sub-set of reports has enabled a detailed focus on food aid and emergency agriculture. The WFP reports are of particular importance as they are thematic in focus and perhaps one of the most extensive evaluations of the agency's operation; also because they are, for the most part, of high quality, comprehensive and rigorous.

#### What was the impact of food aid?

**The reports tell an encouraging story of food aid interventions in many countries meeting their primary objective of feeding the hungry.** Box 3.1 provides some representative quotes (see also WFP, December 2001a; WFP, April 2002; WFP, September 2002; World Vision, June 2001; ECHO, December 2001g; and Handicap International, July 2002 for emergency agriculture).

**This generally positive view that food assistance is saving lives and, in a significant majority of cases, maintaining or improving nutritional status can be confirmed over the three years of the *Annual Review*.**<sup>7</sup> However, the more disaggregated analysis below suggests that despite this overall achievement there are still many improvements that could be made. In particular, evidence to support conclusions regarding results may not always be fully available, and the

targeted population may not be the most vulnerable and may be a fairly limited group given overall needs in the country. Quotes from evaluation reports as to the overall success of the interventions should be read in this light.<sup>8</sup>

#### Were interventions appropriate?

**The reports note varying appropriateness of food aid packages.** The evaluation of WFP's intervention in Angola (April 2002:15) comments: 'From the beneficiaries' perspective, food rations are generally considered adequate and sufficient when the full basket is distributed.' On the other hand, WFP's report on Uganda (December 2001a:9) concluded that: 'While ration levels seem to have been

### Box 3.1 Food Aid Feeds the Hungry

'CARITAS, FALU, WFP and UNICEF monitoring reports indicate a stabilization in health and nutritional status of the population. Starvation has been prevented ... Observations from monitoring and discussions with staff of child institutions indicate that where CARITAS' assistance is able to provide a continuity of supply such as to provide a balanced diet – along with WFP food – the nutritional status of children is much improved from when those supplies are not available.' (CARITAS, January 2002:27)

'WFP has been extremely successful in resourcing the food requirements and delivering them in a timely manner to the camps. The introduction of a new distribution system, scooping directly to heads of families rather than to group leaders, has been a significant achievement in helping to ensure that food delivered is actually reaching beneficiaries.' (WFP, December 2001:9)

'It was considered by WFP that despite the problems of delivery, many lives and assets were saved and that there was a reduction of the numbers of people moving in search of food ... The Evaluation Team gained a strong impression of an imaginative programme approaching major constraints with good judgment.' (WFP, January 2002:10-11)

'The impact of relief efforts in preventing the prevalence of malnutrition from increasing beyond regional norms is worth noting. Over the past two years, nutritional surveys have shown an observable and declining trend in global malnutrition rates.' (ECHO, December 2001c:17)

appropriate given apparent outcomes, setting rations below the food needs of “the most food insecure” may not have been appropriate.’ WFP’s report on Ethiopia (December 2001) notes mixed performance, with general rations in some camps being deficient in protein and micronutrients and the provision of food aid in some areas being above the WFP/UNHCR recommendations. In comparison, the WFP Great Lakes report (September 2002) was positive concerning the provision of micronutrients. The WFP Somalia report (January 2002) includes a detailed discussion of the level of calories provided to pastoralists, noting that a reduction in the ration was based on the faulty assumption that poorer pastoralists had access to alternative sources of oil. Finally, the WFP Ethiopia and Angola reports (December 2001 and April 2002) note considerable trading of food received, suggesting the ration itself was less appropriate than local foodstuffs.

Reasons for providing adequate and appropriate rations in the WFP case included: no pipeline breaks, adequate funds, and cultural sensitivity. The DEC (December 2001) report on post-earthquake assistance in Gujarat found from its extensive community consultation that primary stakeholders were more satisfied with both the quantity and quality of food aid than aid received in other sectors. And the CARITAS DPRK (January 2002) and Oxfam (September 2001) evaluations also found general satisfaction among primary stakeholders with the food assistance provided.

### Procurement

Several reports include an assessment of procurement modes. This is discussed in relation to food assistance in this section and in relation to other sectors in Section 3.4.

The major mode of food procurement is through international donation. A number of questions are raised concerning this, mainly in relation to whether purchasing food locally rather than providing food aid was the best solution. The evaluation of the WFP operation in Somalia includes a detailed discussion of the latter (January 2002:14):

A principal [argument] is that, given the cost of food importation, it would make more sense to give people money to buy local food, whether as cash-for-relief or cash-for-work. It has to be said that in Somalia, as elsewhere, this argument has never been substantially tested by donors/agencies: at some level there is a greater reluctance to fund the

distribution of money in this way than to fund the distribution of food, even at high transport costs.

Other evaluations report different results with local purchases. These are largely dependent on context, in particular the location of markets, and the extent of local production:

- In the Great Lakes, WFP purchased about US\$8m worth of commodities locally. This was viewed as providing a good alternative to commodities such as split peas or yellow maize which were disliked by the refugees, and to bolster the local economy – especially relevant in Tanzania where resentment has grown among the indigenous population in refugee-affected areas (WFP, September 2002).
- The evaluation of WFP's intervention in Ethiopia found widescale trading of food and concluded that: '[T]he economic efficiency of delivering wheat to distant camps so that it can be exchanged for less costly, locally produced cereals is questionable' (WFP, December 2001:9).
- In Somalia, WFP made some attempts to set up local purchase arrangements. However, it encountered considerable problems in securing a trustworthy source of bulk commercial grain at reasonable prices (WFP, January 2002).
- The WFP Uganda evaluation reported that local purchase was conducted on a significant scale and there was a specific programme to build the capacity of small farmer associations that participate in WFP tender processes. During the first year of the PRRO, 33 per cent of cereals and 37 per cent of pulses were procured within Uganda (WFP, December 2001a).
- The FAO thematic review noted that the lack of authority delegated to the field for the purchase of emergency input supplies contributed to considerable delays. A US\$25,000 limit on local procurement – less than the input supplies budget of virtually any project – meant that most procurement had to be carried out by HQ. Delays were caused by correspondence and the need for the field office to identify suppliers. The report concluded that local tendering with an increased limit for emergency activities may be a more cost-effective approach, offering better opportunities for speed (FAO, September 2002).

- The evaluation of World Vision's programme in Kenya commented on a long delay in food assistance coming from Canada and concluded: 'Another major factor contributing to the delayed response was the significant delay in accessing and delivering the therapeutic milk products. ...The time from request to delivery should not exceed two weeks [it was four months]. In order to meet this target, it may be necessary to source products from non-Canadian suppliers and to eliminate the tendering process' (WFP, June 2001:20).

**The evaluation reports reveal that local procurement will only be possible in some cases. At the policy level, however, there is clearly a need to be proactive in ensuring that, wherever politically feasible, food is purchased locally – both for lower costs and to promote local agriculture.**

This is addressed, for example, in the *Euronaid Code of Conduct on Food Aid and Food Security* (1995) which notes: 'In many countries or regions, pockets of need coexist with surplus areas. In these situations we promote the use of locally produced and processed food because it contributes to the development of local markets, reduces costs, improves timing and provides the type of food people are accustomed to. Whenever possible, these purchases should be made from local producers' organizations, thereby promoting their access to the market.'<sup>9</sup> The political feasibility of providing alternatives to food will also depend on the willingness of major food aid providers to either allow food aid to be transferred into cash or to provide direct financial donations instead.

Related to this is the potential for needs assessments, particularly with long-standing refugee populations, to include consultation with women and men concerning their own perspectives on the mode of distribution. That is, would primary stakeholders prefer cash? As the WFP Somalia report (January 2002) notes, this has never been substantially 'tested' by donors and governments. This more detailed needs assessment should also include gender analysis as to the likely effect on gender equality of different kinds of distribution mechanisms. **This is an area that is ripe for investigation with primary stakeholders.**

Peppiatt *et al* (2001) argue that Sen's entitlement theory (lack of access rather than lack of availability) is now widely accepted, but that this has not led to significant attention to cash transfers.<sup>10</sup> However, these authors argue that there appears to be an increasing willingness to consider the use of cash as an alternative to food aid.

Advantages may be the potential for faster delivery and lower transaction costs, injection of cash into local markets, and greater flexibility for primary stakeholders. Disadvantages may be diversion of resources, market distortion, the promotion of inflation, and security risks during transportation and distribution.

### Was coverage adequate?

Several reports, in particular those by WFP, include a strong and nuanced discussion of coverage, both geographical and in terms of targeting of particular groups. **However, the overall finding from this discussion is that too little data is available to be able to make conclusions about coverage results, and many vulnerable groups may be by-passed.**

Part of the reason for this is, in some cases (for example, Angola and Uganda) inadequate donor funding although funding is assessed as adequate in a number of other cases where WFP is active (Somalia, Azerbaijan, Ethiopia). Under-resourcing may occur in some cases because of political decisions of host governments, donors and WFP itself, but there is very limited information on this in the evaluations. This is because the reports tend to steer well clear of wider political issues such as diversion of food aid from one emergency to another, or donors' refusal to fund particular governments. Pipeline breaks, apparently mainly due to mismanagement in WFP (e.g. in Uganda) may be another reason for lack of coverage.

A good example of attention to coverage is the WFP (September 2002b) report on Afghan and Iraqi refugees in Iran. This analyses coverage of the different refugee populations: those inside and outside camps; the differences between camps; the differential needs of those within camps. It notes the general lack of data on those outside camps, and this in turn is linked to vulnerability issues: 'It may be generally true that the more vulnerable refugees are in camps but in the absence of survey data it is not possible to be sure that this is the case or that all refugees outside camps are coping. The camps might have originally accommodated the most vulnerable, but given that the government seems to have more or less frozen the number of refugees in camps it is difficult to maintain that they continue to serve the function of accommodating those unable to achieve self-reliance' (ibid:4–5).

Likewise, the WFP (December 2001) report on Ethiopia provides a similar example of good practice in disaggregation of primary stakeholders. It lists 10 different kinds of refugee whom it interviewed, and notes the potential for repatriation and

establishing livelihoods for each category. This level of analysis should be extremely useful in terms of targeting assistance, providing that adequate records can be maintained.

Coverage issues are highly political. For example, the WFP Great Lakes (September 2002) evaluation explains that one indicator of a successful program is that refugee populations demonstrate consistently better nutritional status than the host population. Does this mean that other vulnerable groups are being missed by government and aid agencies? Extensive attention to refugee camps as opposed to IDPs was also raised in *Annual Review 2002* as a highly problematic area. Food aid objectives often include a relatively narrow section of the population and evaluators need to assess whether such objectives are relevant to need.

### Box 3.2 Reaching the Most Vulnerable Through Food Aid

'IDPs are probably the most disadvantaged group living in the area. They are not targeted by the PRRO, but are entitled to request food from refugee kinsfolk. In consequence the ration provided by refugees is used to feed a larger number of people than intended by WFP.' (WFP, December 2001:6)

'WFP conducts geographical targeting in so far as the standard rations provided to IDPs and refugees vary depending on the district and the camp or settlement ... There is however no targeting within camps/settlements ... The result must be that the "most food insecure" (who typically represent about 30% of the camp/settlement population) receive only a portion of what they need to bridge their (greater) food gap.' (WFP, December 2001a:28)

'Within refugee camp situations, although it has been possible to distinguish different income groups, there are no practical guidelines for selecting the most vulnerable households for higher generalized rations. There are also no standard methods for modifying targeting criterion at times of resource shortfall. Decisions tend to be made on a purely ad hoc basis.' (WFP, September 2002:20)

'Beneficiary identification was found to be weak in nearly all interventions ... only rarely was there a poverty focus in the beneficiary selection process. This is not surprising since, in many cases, those responsible for compiling beneficiary lists do not have the skills or time to determine who are the most deserving recipients.' (FAO, September 2002:31)



**There are a number of examples of situations where either the needs of the most vulnerable have not been met or are not understood** (see Box 3.2 for representative quotes; for similar points see ECHO, December 2001f). Needs assessments were assessed as requiring improvement in the majority of cases (e.g. WFP, December 2001, September 2002b, January 2002; World Vision, June 2001; FAO, September 2002).

The issue of whether it is appropriate to target the most vulnerable is also raised. The conclusion here, supported by findings from the two previous *Annual Reviews*, is:

- 1 Food aid providers need to determine whether it is culturally, politically and socially appropriate to be targeting the most vulnerable.
- 2 If it is culturally appropriate, greater attention needs to be paid to understanding who the most vulnerable are as well as the most effective ways of targeting these groups.

Although cultural appropriateness and intra-community distribution is not covered by the evaluation reports in similar depth as those analysed for *Annual Review 2002* (p135; six reports), the issue is covered in four reports. The CARITAS DPRK evaluation (January 2002) comments in the context of aid diversion that ‘the question of aid not reaching intended beneficiaries as intended appears to be more a question of assistance sharing – among families and communities. While the former is more or less accepted by international agencies, the latter is more problematic as both usage and impact are much harder to assess.’ The WFP Great Lakes evaluation (September 2002:20) notes:

Targeting, based on livelihood criteria or personal and household assets, is very complicated and problematic to implement properly even with good guidelines ... Ensuring that women or vulnerable individuals (usually individuals with low status) receive the rations does not necessarily ensure that they will retain control over that food resource within the household ... Targeting vulnerable individuals in a household or community in a culture where sharing is the norm can actually put that individual at personal risk when there is a shortage of food and/or other resources.

The World Vision/UNICEF evaluation of their nutrition intervention in Afghanistan (World Vision, September 2002:9) includes in-depth discussion on this issue:

One clear and consistent finding from interviews with community members was that the SF [supplementary feeding] rations were certainly shared amongst the entire household. In some cases, the food was also shared with neighbours who did not qualify for the program. Thus a distribution that was intended to last two months lasted only two weeks in many households, with the targeted beneficiaries receiving only a fraction of their intended ration ... A common complaint in community interviews was that the programme should have provided food to all households i.e. blanket rather than targeted distribution.<sup>11</sup>

**Intra-community distribution mechanisms are at odds with agency policy,** an area that should be covered more systematically by evaluations in their assessment of coverage. In some cases the focus on the most vulnerable should be on whether the nutritional and overall status of this group is improving, rather than merely attempting to target this group with aid. Intra-household distribution mechanisms are also not well understood. However, because of the significant time and resources required for analysis of the implications of this, this will require attention in both monitoring and evaluation efforts. At the same time the policies in place in most agencies to promote support to the most vulnerable may be in contradiction with the aim of promoting LRRD, as it is often the most vulnerable who have most difficulty moving permanently beyond the relief 'phase'.

#### Connectedness in food aid

WFP's seven evaluations of PRRO interventions, as well as the FAO's thematic review, provide the opportunity to examine connectedness in unusual detail. This has implications beyond the two agencies considered.

The PRRO programming modality was introduced by WFP in April 1998 after the endorsement of the policy paper *From Crisis to Recovery* (WFP, 2000). PRROs run for three years and are expected to include an appropriate mix of relief and recovery activities; funding is transferred between these two modalities by the county director as the situation requires. As such, the PRRO is intended to act as a bridge from relief to development, and to include planning for a more comprehensive recovery or development programme as well as for exiting or down-scaling relief initiatives. The policy paper responded to pressure on relief organisations to show that their work was likely to promote sustainability in addition to WFP's need for a more secure funding base. How did the PRROs work in practice in terms of promoting connectedness and supporting sustainability?

The evaluations found that the PRRO category may have continuing relevance at the policy level in that agencies involved in relief activities should continue to take longer term perspectives into account. For example, the Angola evaluation (April 2002) notes that it was a major achievement for WFP to have defined a recovery strategy in a vacuum as neither the government nor the UN had defined a recovery strategy under which WFP could work. However, the majority of interventions focused mainly on relief and, while they were largely successful in feeding hungry populations, did not make the necessary linkages to recovery and development. Despite there being a policy in place, LRRD has once again proven problematic.

The Great Lakes report is representative of findings as a whole (September 2002:vi):

The complex regional environment is not conducive to recovery activities. Constraining factors identified ... include national policies that do not favour some recovery activities [and] United Nations and government security restrictions have continued to place serious limitations on programming opportunities beyond the provision of immediate relief. Another factor has been the reluctance of some WFP staff to serve in the difficult and unpredictable duty stations ... Many of the recovery activities observed by the mission in Rwanda have been heavily biased towards rural infrastructure improvements, without a strong focus on building lasting assets for beneficiaries[.]

This picture is repeated in several other reports, for example WFP's Ethiopia report (December 2001:24): '[T]he result is an operation that has remained focused for the most part on care and maintenance, while recovery elements, contributing to building sustainable livelihoods for the repatriated and dispersed and the encouragement of self-reliance in the camps, received less attention and support. Potential opportunities for recovery may have been lost.' In contrast the WFP Somalia report (January 2002) notes that funds could be switched back and forth between relief and rehabilitation as the situation required.

There are a number of interconnecting reasons for the inability to move beyond relief:

- The security situation has not allowed substantial attention to recovery in some cases, although security as an issue is highly variable (e.g. WFP in Angola, Great Lakes, Uganda, Somalia).

- It is often difficult to plan for longer than the period of a year, which is one of the time constraints of the CAP. Donors may not be willing to provide funding for more than one year.
- Government policies do not often support recovery (e.g. such as building self-reliance among refugees in Tanzania); and there is a lack of government capacity to take over from external agencies and promote recovery.
- There is lack of capacity in the WFP and FAO country offices (e.g. WFP offices in the Great Lakes and Iran).
- There is lack of capacity in implementing partners (e.g. WFP in Somalia), and absence of policies/strategies on how to support implementing partners.
- There is lack of synergy between emergency and development departments in agencies.

LRRD policies may be encouraging agencies to set unrealistic objectives in terms of recovery ('results inflation'). For example, a five-country study of recovery after major natural disasters has found that many poor and vulnerable people may only temporarily move out of the relief phase (ProVention Consortium, 2002). The implications are that a more sophisticated typology of recovery situations is needed combined with key indicators of likely success, such as the existence of geo-political stability, strong implementing partners, and local and national capacities including government capacity.

## Water and Sanitation

### 3.2.2

This section is based on eight reports: five NGO reports, by the DEC, CARE, Tearfund, and Oxfam (two reports); ECHO (two reports); and one UNICEF report. The countries covered are: India, Afghanistan, Angola, Burundi, Sierra Leone and the Sudan. The sample is somewhat limited as several of the reports (DEC, CARE, Tearfund and UNICEF) consider water and sanitation as part of a wider sectoral analysis, and several of the interventions are relatively small scale. Findings in this sector are, however, similar to those for *Annual Review 2002*. Conclusions have been drawn from across the two years.

### What was the impact of water and sanitation interventions?

The evaluations agree that interventions were successful in meeting physical targets and output-related objectives. Agencies have been effective in installing handpumps and protecting water sources as well as building latrines. Representative quotes and examples are included in Box 3.3.

*Annual Review 2002* noted similar successes in building physical structures and reducing mortality and morbidity. Some of the reasons for success are provided in ECHO's report on Sierra Leone (December 2001c:14) which notes that: 'Experience elsewhere has demonstrated that a comprehensive approach can influence the health status of a community and be more effective than establishing

#### Box 3.3 Success in Water and Sanitation Interventions

'Oxfam GB's public health programme has responded to the considerable needs of the displaced population in Angola with a high degree of commitment and appropriate action. The programme has a strong hygiene component resulting, in many of the programme areas, in a high level of understanding of key hygiene behaviours. Over 1200 water points provide potable water and more than three thousand latrines ensure adequate sanitation. Coverage and quality are generally good. Despite the considerable challenges presented by the context, internal recruitment and management problems, Oxfam has delivered a programme of real value to the beneficiaries.' (Oxfam, June 2002:1)

'The ECHO/UNICEF Drought Emergency Project in Sudan has played a key role in reducing ... drought effects[.] Where no interventions took place people remained camped in their areas of migration and continued to put pressure on land and existing water facilities.' (UNICEF, December 2001:22)

'CARE's continued operation of the Kabul water project throughout the bombing probably saved lives.' (CARE, September 2002:47)

'The situation of assisted populations has clearly been improved in areas where water and sanitation projects have taken place. Improved conditions and behaviours are reflected in a lower incidence of disease[;] morbidity and mortality due to bloody diarrhoea were controlled and no increases were seen during the period under review.' (ECHO, December 2001c:17)

the infrastructure only. Over the past year ECHO-funded projects have gone beyond building wells, latrines and installation of hardware. A combination of approaches including hygiene promotion, safer disposal, hand washing and maintaining drinking water free of contamination has been adopted by ACF, GOAL and OXFAM.'

However, two of the problem issues highlighted in *Annual Review 2002* are also found in some of this year's reports: **limited primary stakeholder participation in planning and design, and lack of operation and maintenance capacity**. The fact that the latter issue has surfaced in 15 reports included in this and last year's *Annual Review* suggests that it is a systemic problem.

### Consultation and Participation

Consultation and participation may not always be feasible dependent on the phase and intensity of the emergency. The interventions considered in this section are either recovery focused or linked to recovery, and in principle could have facilitated primary stakeholder consultation and participation.

Two reports (Tearfund, July 2002; Oxfam, March 2002) note a general lack of participation. The Oxfam report concludes that the low degree of primary stakeholder participation probably led to a proportion of the population in the intervention target area being too far from protected springs to benefit.

On the other hand the evaluation of ECHO's intervention in Sierra Leone concluded (December 2001c:14): 'In the IDP camps, there was a growing realisation that a demand-driven approach through increasing people's participation contributed to coverage and use more than a supply-driven approach. The projects in rural communities have involved village members in implementation (e.g. digging wells and pits and constructing structures), health education and monitoring. Flexibility in the choice of technical options that suit different groups within a given area has been ensured, as reflected in the latrine design for school children in the GOAL-funded project.' Unfortunately no information is provided in this report as to how this increased participation was achieved.

Three reports also found evidence of gender balance in operation and maintenance committees, but otherwise limited participation of women (Oxfam, March 2002 and June 2002; ECHO, December 2001e). The Oxfam Burundi report mentions in

passing (March 2002:3) '[The programme's] gender strategy consisted mainly of ensuring near-equal gender representation on hygiene and water committees. Had the project been more participatory in the identification, design and monitoring phases, there would have been more opportunity to incorporate further gender-specific input, analyses and objectives.' More detail is provided in the Oxfam Angola report (June 2002:21). In this case, while it is noted that there is gender balance in membership of water committees (*Grupos de Agua y Saneamiento*), most of the chairpersons and treasurers were men, there were no female technicians, men tended to dominate in community meetings, and Oxfam mobilisers allowed this to happen. While such gender-based issues are fairly common in the water sector, EHA could pay more attention to both the issues and means to overcoming them.

### Connectedness in water and sanitation interventions

Findings of the reports on participatory management mechanisms for water facilities in particular point to poor results as well as a lack of institutional learning. Poor results are noted in a number of reports (and see Box 3.4):

- In the UNICEF (December 2001) intervention in Sudan, it was noted that only about 65 per cent of all the hand pump installations in the three states covered were operational. The report recommends that the state's water authorities stop any drilling and installation of new hand pumps and concentrate on rehabilitation of the old ones.
- Tearfund (July 2002) comments on the use of non-local parts which was likely to cause problems.
- In Angola, the future of water committees in Oxfam projects, which had been trained to repair and maintain hand pumps, was considered uncertain because of confusion among implementing agencies concerning what their future role should be (Oxfam, June 2002).
- The evaluation of ECHO's intervention in Sierra Leone (December 2001c) comments on hand pumps which have fallen into disrepair due to lack of maintenance.
- The ECHO evaluation in Burundi (December 2001e) notes weakness in mechanisms designed to ensure the maintenance of infrastructure, and the

minimal importance given to water quality surveillance and the actions directly related to it.

- The DEC report on Gujarat (December 2001) notes that much time and DEC resources have been wasted on the unsuccessful promotion of latrines: 'Considerable long-term investment in health promotion will also be needed to ensure that the latrines included in some of the DEC-funded housing schemes will ever be used, and to prevent them becoming a health hazard in themselves. Few lessons seem to have been learned from the Orissa Cyclone, where Oxfam also ran into difficulties trying to implement a large sanitation programme' (Vol. 3, Report 4:1.5.3). Part of this issue relates to cultural sensitivity and the lack of understanding of local context.

The consistent attempt and partial failure to develop committees for water management reveals a lack of learning in the sector; a cursory search also found little guidance in this area for emergency or near emergency situations even though it is a major part of water and sanitation initiatives.

#### **Box 3.4 Sustainability of Water Facilities**

Oxfam's intervention in Burundi, funded by ECHO, achieved its stated aims in terms of the numbers of springs protected and sanitary facilities provided; it also contributed to reduced morbidity and mortality. The plan for this intervention was for most of the work to be carried out through water and hygiene committees, which were formed and trained for this purpose. However, the evaluation (Oxfam, March 2002) notes a number of problems:

- the project set no indicators or long-term goals regarding sustainability of the committees;
- the project may not have researched and tried to tie in to pre-existing community structures;
- the committees did not have clear responsibilities or a timeline for operation.

One possible solution offered by the report to these generic problems is discussion with communities in the project planning phase as to the nature, make-up and mechanisms to maintain the sustainability of such committees.



One illustrative example of the lack of learning and lack of concomitant use of evaluation findings over time is Oxfam's intervention in Angola (June 2002). An ECHO evaluation (ECHO, 2001:13) which took place in November and December 2000 and which covered Oxfam's intervention pointed to some of the shortcomings with attempting to establish sustainable solutions: 'A lasting impact on the water supply is not guaranteed, since the project did not develop a long-term pump maintenance strategy'. By June 2002, when the Oxfam evaluation reviewed for this section was published, serious problems remained with its strategy for sustainability.

## Health

### 3.2.3

The following is based on five reports which focus specifically on health (Oxfam, July 2002; ECHO, October 2001a; ECHO, October 2001c; Norway Ministry of Foreign Affairs, November 2001; and MSF, November 2001), and five reports which include a discussion of health issues as part of a wider focus (Oxfam, June 2002; DEC, December 2001; ECHO, December 2001c; ECHO, December 2001g; ECHO, February 2002). Several interventions also included a health education component.

#### What was the impact of health interventions?

**As found in *Annual Review 2002*, interventions were seen to have largely met their health-related objectives – usually stated in terms of stabilising mortality and morbidity rates and ensuring that infectious diseases were kept under control.** Synthesising the findings from health-related projects in five countries, the Norwegian Red Cross (NRC) report found that (Norway Ministry of Foreign Affairs, 2002:29): 'All the projects visited by the evaluation team showed a high level of correspondence between the objectives and the timely achievement of outputs, as well as between the reporting on outputs and the outputs that could be verified. The international operations show a uniformly high level of effectiveness as regards output delivery in a wide variety of environments... [R]eliable funding and the strong work ethic of its personnel have often been key to the NRC's ability to meet its objectives.'

The MSF evaluation in El Salvador concurs, with some qualifications (MSF, November 2001:25): '[T]he mission succeeded in implementing the right activities

### Box 3.5 Success in Psycho-Social Interventions

Four reports evaluate psycho-social interventions as part of wider health interventions:

- The evaluation of MSF's programme in El Salvador (November 2001) notes that this was a significant feature of its intervention, and very successful. It involved a short intervention of eight weeks giving support in the immediate aftermath of the earthquake, with a total of 1,815 clients receiving individual support and 859 receiving support in groups.
- The evaluation of DEC-supported activities after the Gujarat earthquake found (December 2001:31): 'ActionAid viewed its psycho-social work within a social context and linked it to public information and advocacy work which brought practical benefits to people in distress. As such it was an example of good practice in an area where there is a real risk of cultural imposition.'
- The ECHO evaluation of child protection activities in Sierra Leone also considered the psycho-social elements of support as generally successful, despite reservations about long-term impact (December 2001b:20): 'The most important impact is that the children are separated from the adult combatants and are provided with a safe environment, given appropriate medical-psychological-social assistance, respecting his/her culture and religion. The child gets education again ... In general, there is compre-

in an efficient manner, for the correct duration of time. Against this, the failure to provide a coherent operations plan led to a partial failure to explicitly target the beneficiaries most in need.' The DEC evaluation (December 2001:41) notes: 'A number of agencies provided good responses in terms of curative care... [but] the attempt to increase community capacity through health education programmes was less effective.'

On the other hand several highlight weak results-based planning. This makes it difficult to draw conclusions due to both poor monitoring practices, lack of data, and/or inappropriate gathering of data. In the case of the Oxfam intervention in

**Box 3.5 Continued**

hension and respect for the child's suffering. An ex-combatant is perceived more as a victim than a perpetrator ... It is too early to measure the long-term impact of the programmes. Above all, an assessment is needed of the quality of the integration of children in their family and the community.'

- The evaluation of ECHO's support to IDPs in Colombia (February 2002) also notes that psycho-social interventions have been successful in bridging the gap between relief and development by strengthening the coping capacity of IDPs to respond to military invasion of their lives.

Activities that may have a psycho-social element are not included under that heading. For example, the evaluation of assistance to returnees and vulnerable groups supported by the Netherlands Red Cross (Netherlands Ministry of Foreign Affairs, January 2002) describes in detail the importance for returnees of their ongoing interaction with Red Cross volunteers. Such interaction is often not picked up in monitoring and evaluation, especially when focus is on numerical outputs. As the ECHO evaluation in Sierra Leone points out (December 2001b:20) 'For the partners, it is clear that the number of children who come through the ICC [Interim Care Centres] is not the only factor – the psychological state of the children is also important. Partners emphasise that a single child can mobilise an agent for several days or even weeks, whereas four other children will need only relatively little support.'

Angola (June 2002) too much baseline data was collected and what was collected was not analysed or used. Further, ECHO (December 2001c:17) comments: 'As a proper surveillance mechanism is lacking, the impact of immunisation coverage is not yet properly documented. Despite the investment, the programme might not result in the expected reduction of morbidity and mortality. While sporadic cases of measles and tetanus have been reported by medical NGOs, the achievements in tetanus immunisation coverage and child immunisation remain to be seen.'

Psychosocial interventions were also found to have generally met their objectives and were evaluated as successful (Box 3.5).

### Connectedness in health interventions

As in the water and sanitation sector, establishing physical structures has proven easier than ensuring that these facilities are used and maintained. Several reports focus on the sustainability of government health structures and the key finding from this year's reports, supported by three reports from *Annual Review 2002*, is that **while agencies are effective in meeting short-term health needs there is a failure to link these short-term interventions to longer term planning.**

A case in point is provided by the NRC's evaluation (Norway Ministry of Foreign Affairs, November 2001:30):

The most significant problems faced by the NRC in achieving satisfactory outcomes are illustrated in the rehabilitation of the Nyanza hospital in Rwanda ... The evaluation visit to the hospital showed a very low attendance rate (less than a third of capacity for the total hospital and a quarter for the surgical unit), and understaffing. Even though the structure itself was of a high quality and hence relatively stable in itself, there was evidence that the NRC investment had ignored contextual factors that dictated the relevance of the project.

These contextual factors included: the price of health services; difficulties of transportation; distrust in the quality of the staff; ethnic distrust, with patients coming from one ethnic background and healthcare providers from another.

This report also notes problems in relation to handing over institutions and work to national Red Cross societies. The reasons given for this are that the assistance standards set by the NRC are often beyond the reach of the local Red Cross and most local actors. The report therefore recommends (Norway Ministry of Foreign Affairs, November 2001:32): '[T]here is a need for a more extensive analysis of the ways in which resources can be transferred to local actors ... The exit strategy and handover procedures should be integrated into the project's contextual analysis from the start and focused on throughout projects, not only towards the end.'

A detailed analysis is also found in the ECHO evaluation of its health, nutrition and water and sanitation interventions in Sierra Leone (ECHO, December 2001c:ii-iii):

The contribution that projects have made to restore basic health services has created better conditions for health staff... [but] the overall issue of

long-term capacity of public services to sustain a certain level of delivery remains a very significant constraint ... From a project perspective, it may be argued that financial sustainability can be assured through the mobilisation of extra resources through cost recovery. From a sectoral perspective, however, the scale of the problem becomes apparent. The volume of recurrent support being provided by ECHO partners cannot be absorbed by the MoHS or addressed through user charges within the foreseeable future. Thus the viability of the 'start-up-and-hand-over' paradigm is called into question.

This report recommends a better needs assessment process whereby development of health activities is based on a comprehensive plan, developed in association with local actors, and initiatives such as improved management, supervision and training are prioritised. As noted in the quote above, the evaluation of the NRC (Norway Ministry of Foreign Affairs, November 2001:51) also recommends paying closer attention to context during planning, arguing that: 'The Government [of Norway] should require that the relief projects carry out more monitoring and analysis of the social, cultural and institutional factors of success of projects in harmony with current practice in the Development Section.' Analysis of a similar kind can also be found in Oxfam (July 2002), DEC (December 2001), ECHO (October 2001a), and ECHO (October 2001c).

As with the water sector, it may be that there are no clear standards or guidelines in most agencies that deal with sustainability and how local capacity can be supported. **As lack of capacity in relation to understanding local context and planning on the basis of this understanding has been highlighted in several reports, it is important for agencies to focus more attention on this, with specific reference to support for local capacity.**

## Housing and Temporary Shelter

### 3.2.4

Four reports evaluated housing interventions – considerably less than the 10 reports analysed in *Annual Review 2002*. These housing projects were smallscale and should not be considered representative of housing interventions as a whole. For example, in the Gujarat case some one million houses are being rebuilt or reconstructed; DEC-supported NGOs are involved in probably less than 20,000 units. Temporary shelter-related activities were covered only in the DEC report in sufficient detail for

inclusion here.<sup>12</sup> Overall the results of housing interventions support the findings of *Annual Reviews 2002* and *2001* concerning the inability of relief interventions to properly support housing reconstruction. Major problems reported in all three years include lack of primary stakeholder consultation and participation, inappropriate design and siting, and high cost.

After the Gujarat earthquake in January 2001, the DEC supported interventions by NGOs who aimed to provide shelter and more permanent accommodation to the homeless. Good practice from this intervention in construction of temporary shelter is highlighted in Box 3.6. In the housing sector it was found that (DEC, December 2001):

- there was limited consultation on the design and size of houses;
- contractors rather than communities were the principal builders;
- ‘partnerships’ were unidimensional, with plans, processes and products mainly shaped by urban consultants often unfamiliar with rural conditions;
- innovation in design, technology and organisation was limited;
- costs, both construction and organisational, were high;
- new houses and villages leave much to be desired in appropriateness of design, construction quality, primary stakeholder satisfaction, employment opportunities, and links to longer term development.

The report also discusses the failure to recognise the multidimensional nature of primary stakeholder livelihoods (DEC, December 2001, Vol. 2:13): ‘An important characteristic of the area is that for a majority of people houses are workplaces as well as homes, and so the earthquake destroyed livelihoods as well as lives. The problem was not simply the direct destruction of workplace homes but also that people could not go out to work because they had no safe place to keep their possessions. The rapid provision of semi-permanent, secure structures was extremely important for recovery. SCF and Caritas (CAFOD) were the only agencies to recognize this issue. This is a serious shortcoming of the DEC response, especially because funds were lying unused at the crucial time.’

A Tearfund (April 2002) supported intervention, aimed at improving the living conditions of 850 IDP/refugee households through housing improvement in collective centres (CCs) and private accommodation (PA), found that:

- PA but not CC interventions generally met Sphere standards;

- beneficiaries were fully consulted and were generally satisfied with the response;
- others in the community who did not receive assistance gave the intervention a mixed reception;
- the project provides a two-year rent free period for the beneficiaries. This enables them to re-establish their personal lives. However, over the longer term the owner of the building may turn out to be the main beneficiary, given improvements made. Furthermore, the intervention is only a temporary solution to the issues faced by IDPs/refugees.

The evaluation (Oxfam, May 2002) of an Oxfam-supported intervention providing 157 houses to flood victims, and funded by ECHO, was generally negative:

- planning for the programme was poor and substantially overlapped with the government programme, which could have provided houses supported by Oxfam, even if to a lower quality;
- the project was rehabilitation-oriented, verging on urban development;
- beneficiary participation was successful as far as construction was concerned, but needed significant input from Oxfam;
- site location and housing design were appropriate, but because of fortunate circumstances rather than Oxfam design;
- site layout was well planned, but serious constraints were being faced at the time of the evaluation in provision of sanitation services;
- coverage may have been problematic in terms of exclusion (ibid:25): ‘In targeting only those affected by the flood with one specific intervention, the Pernambuco project addressed a very few of those experiencing asset depletion and long-term livelihood decline.’

A further Oxfam (September 2001) programme, this time in Bangladesh, involved the reconstruction of 9,000 houses after floods in 2000. The evaluation of this project concluded:

- coverage was limited as over half-a-million houses were destroyed;
- individual villages were targeted, leaving neighbouring villages excluded;
- design was inappropriate and did not incorporate local suggestions;
- 20-30 per cent of houses were not being used as intended, and some houses were built for households outside of the target criteria;
- land tenure was a major issue.

**Box 3.6 Good Practice in Temporary Shelter**

SCF and its partner agency Abhiyan supported temporary shelter after the Gujarat earthquake in 2001. Their intervention was assessed as follows: 'Over 24,000 units were constructed in 250 villages through 21 partner agencies in six months by Abhiyan, in partnership with SCF. This shows how local NGOs can scale up [and] marshal considerable managerial resources and materials ... The cost was low at Rs.4000 (£60) for a unit of 225 sq. ft and the design was conducive to community-managed construction ... Beneficiaries are now generally satisfied ... A more significant benefit of the participatory method (in which beneficiaries dig foundations, raise plinths, and construct walls) is that it put the shocked communities to work. Investment in reusable materials (tiles and bamboo) indicates judicious use of available resources and a long-range strategy' (DEC, December 2001, Vol. 3, Report 2: pt 18). 'The main lesson from the Latur earthquake in India only eight years previously had been to help people rebuild their own houses in their own way' (DEC, December 2001, Vol. 2: 45).

An important finding from the DEC evaluation and past evaluations reviewed in this *Annual Review* is that self-construction, supported by external agencies, is likely to be the most effective reconstruction technique – as long as the fact that some vulnerable groups may not be able to reconstruct their housing is taken into account.

## Other sectors

### 3.3

As the remaining reports are divided between a large number of sectors the findings are summarised for ease of reference (Table 3.1). Lack of capacity building is also a theme in these examples.



**Table 3.1 Findings from Other Sectors**

| Sector, Agency and Purpose  | Main Findings and Issues  |
|---|---|
| <p><b>Non-food relief</b></p> <p>IFRC (September 2002). Support to 6,700 families with plastic sheeting, jerry cans, kitchen sets, blankets, etc, after Goma volcanic eruption.</p> | <ul style="list-style-type: none"> <li>● The operation was moderately successful.</li> <li>● The response was rapid, substantial and relatively well-coordinated.</li> <li>● The majority of the assistance provided was appropriate.</li> <li>● There was difficulty changing focus to assist returnees in Goma after the unexpected mass return there.</li> <li>● Reporting against Sphere standards was not carried out.</li> </ul>  |
| <p><b>Non-food relief</b></p> <p>CARE (September 2002). Work in several rehabilitation sectors including rebuilding roads, drains, renovating water systems, food-for-work.</p>     | <ul style="list-style-type: none"> <li>● Appropriateness of the assistance was confirmed by interviews with beneficiaries.</li> <li>● Most relief/rehabilitation activities were highly rated.</li> <li>● Renovation of underground water abstraction systems carried out by CARE (<i>Karezes</i>) was consistently identified by primary stakeholders as a rehabilitation priority.</li> <li>● Food-for-work may lead to disincentives to local production; cash-for-work should be considered as an alternative in order to support rural livelihoods.</li> </ul>   |
| <p><b>Mine Action</b></p> <p>UNMAS (February 2002). Mine clearance and education funded by DFID and other donors (see Box 3.11 for good practice in coordination details).</p>      | <ul style="list-style-type: none"> <li>● Overall the programme was very successful: nearly 45,000 lethal devices were destroyed, and over 30 million m<sup>2</sup> of land were restored.</li> <li>● The number of mine-related incidents has dropped dramatically.</li> <li>● Coordination by the UN was largely successful.</li> <li>● The quality of mine awareness programmes was variable.</li> <li>● Handover of the programme to the Kosovo Protection Corps was seen as a political decision by the Kosovo Force (NATO), which undermined attempts to develop local capacity.</li> <li>● The programme received generous funding (some US\$72m), unlikely to be available to other mine action programmes in less visible complex emergencies.</li> </ul> |

**Table 3.1 Findings from Other Sectors** *continued*

| Sector, Agency and Purpose   | Main Findings and Issues   |
|--|--|
| <p><b>Support to IDPs</b></p> <p>ECHO (December 2001a).<br/>A 2.7million Euro intervention supported by ECHO providing support in health, water and sanitation, nutrition and distribution of non-food items; and dissemination of the Red Cross/Red Crescent Code of Conduct.</p> | <ul style="list-style-type: none"> <li>● Response was appropriate and had a positive impact in terms of avoiding the outbreak of water-related diseases.</li> <li>● Provision of non-food items played a useful role in improving the hygiene conditions in camps and also supported the resettlement of several thousand IDPs in their villages of origin.</li> <li>● Malnutrition levels were similar to those observed in non-IDP populations.</li> <li>● Dissemination of the Code of Conduct had overly ambitious objectives which were unlikely to be fulfilled.</li> <li>● Attempts to link emergency and recovery funding have been unsuccessful.</li> </ul>   |
| <p><b>Support to refugees and returnees</b></p> <p>ECHO (2001h).<br/>Evaluation of UNHCR activities supported by ECHO in Serbia, Kosovo, Zambia and Guinea.</p>  | <ul style="list-style-type: none"> <li>● Appropriate protection measures were taken in all cases, although the percentage of the budget allocated to protection was minimal (e.g. 4 per cent in Guinea).</li> <li>● Coordination of major activities in favour of refugees was criticised in Guinea and Zambia but was found to be positive in Serbia. Here UNHCR was instrumental in ensuring the import of goods into the country during the Milosevic regime, as well as protection to its implementing partners.</li> <li>● Direct implementation of programming takes up at least half of UNHCR's budget and is hampered by lack of qualified technical field staff, sudden budget cuts and weakness of implementing partners.</li> <li>● Data on direct implementation is not maintained in a way that supports analysis of cost-effectiveness.</li> </ul> |
| <p><b>Assistance to returnees</b></p> <p>Netherlands Ministry of Foreign Affairs (January 2002).<br/>Assistance to the Croatian Red Cross to support 6,400 primary stakeholders with information, food and welcome parcels and other relief supplies.</p>                          | <ul style="list-style-type: none"> <li>● Mobile teams and Red Cross volunteers effectively delivered packages to primary stakeholders on a regular basis, and had an important protection function through their contact with stakeholders</li> <li>● Stakeholders, often in remote settings, very much appreciated the regular contact with mobile teams as well as the resources provided.</li> <li>● The objective of reducing the reluctance of refugees to return to Croatia may not have been met.</li> </ul>  |

**Table 3.1** Findings from Other Sectors *continued*

| Sector, Agency and Purpose   | Main Findings and Issues   |
|--|--|
| <p><b>Income generation</b></p> <p>CARE (March 2002). EU-funded facility for job creation and promotion of return of refugees in Bosnia and Herzegovina.</p>   | <ul style="list-style-type: none"> <li>● Very high degree of satisfaction of beneficiaries.</li> <li>● Overall economic and social impact considered to be quite high.</li> <li>● Sustainability of job creation is high.</li> <li>● Programme for capacity development of NGOs is less strong.</li> <li>● Programme supported ethnic reintegration.</li> </ul>  |
| <p><b>Income generation</b></p> <p>DRC (April 2002). Five-country review (Azerbaijan, Georgia, Serbia, Montenegro, Somaliland) of income-generating activities (IGA) aiming to promote post-conflict rehabilitation.</p> | <ul style="list-style-type: none"> <li>● Programmes achieved their planned outputs in terms of the number of primary stakeholders.</li> <li>● It is unclear what impact IGA schemes had on primary stakeholders.</li> <li>● As organisational sustainability is an overriding concern in micro-finance it may not be an appropriate intervention for humanitarian actors.</li> <li>● Lack of sustainability may be promoted by the focus of humanitarian actors on targeting the most vulnerable, who may be the least able to become self-sufficient.</li> <li>● Women were given special attention.</li> </ul> |

## Elaboration of Cross-cutting Themes

### 3.4

### Human Resources and Staffing

#### 3.4.1

**One of the central factors in the success of humanitarian action has been the dedication of staff – ordinary people doing extraordinary things, despite working in disabling bureaucracies** (e.g. WFP, January 2002 and, September 2002c; ECHO, December 2001c; Tearfund, July 2002; CARE, March 2002; World Vision, June 2001; Oxfam, May 2002). One of many references (ECHO, December 2001c:15) notes: ‘It is worth noting the excessively long hours of work which expatriate staff are doing in order to meet the activity timeframes. Some are visiting health facilities seven days a week. This reflects the fact, that within

a context of medium-term planning, projects are still implemented in emergency mode.<sup>13</sup>

**As in *Annual Review 2002*, reports this year highlight staff turnover and inappropriate hiring of short-term expatriates as two of the most significant problems in terms of staffing.** Indeed, 14 reports note the detrimental effects of staff turnover on interventions.<sup>14</sup> The Norway Ministry of Foreign Affairs report (November 2001:33) elaborates on this problem: 'The changes in personnel in charge of projects both (sic) at headquarters, in the country capital and the field are not easily justified ... Frequent changes in staff mean that the memory of projects is sometimes forgotten ... The lack of good and systematic reporting further affects the institutional memory and an informed appreciation of projects.'

Just how extreme this situation can become is illustrated by the evaluation of Oxfam's intervention in Angola (Oxfam, June 2002:22): 'In the last year, 32 international staff have filled 11 posts; an average of three incumbents per post per year. A snapshot of the stability of the senior positions during the two week period of the evaluation shows that out of the 13 senior posts, only two ... were fully in post. Eleven were acting, interim, absent due to visa problems, on R&R, or had only just arrived.' High staff turnover among both international and national staff was also described (CARE, March 2002), in this case because of low salaries; the WFP synthesis report on its *Commitments to Women* found higher turnover of female than male staff (September 2002a).<sup>15</sup>

This problem is compounded by what is seen as an inappropriate appointment of international staff (see Box 3.7). The DEC Gujarat evaluation (December 2001, Volume 3, Report 3:3.4.2) discusses this issue in some detail:

Expat staff are a very great deal more expensive than India staff. Salaries of managers are four to eight times higher, and support costs are also high (including international airfares, rest and recuperation, and UK based recruitment costs). Many Indian managers demonstrated project management skills that were as good as or better than those of expatriates ... expats who are not familiar with India have to learn much about the local context, and are more likely to take inappropriate decisions ... [S]ome member agencies found it difficult to attract well-qualified Indian staff,

particularly as they were only able to offer short-term employment contracts.

The evaluation of CARE's programme in Afghanistan (September 2002:28) similarly points out: 'The effective utilisation of experienced national staff typically led to significantly lower staff costs than would have resulted from the use of expatriates.'

Other problems were evaluated in relation to management, some of which may be linked to staff profiles:

### Box 3.7 Expatriates to the Rescue?

'Too many FACT [Federation Field Assessment and Coordination Team] and PNS [Participating National Society] delegates were deployed, with inadequate mission instructions and, in some cases, inappropriate skills. Some FACT members stayed longer than required. The RRC [Rwandan Red Cross] felt overrun by delegates.' (IFRC, September 2002:1)

'Much of the initial reporting about the earthquake focused on the British "search-and-rescue" team. While thousands of people were extracted from the rubble by neighbours and government staff, the UK media focused on the 6-strong British team which rescued just 7 people.' (DEC, December 2001:44)

'In practice, most expatriates are not only responsible for major decision-making but also for day-to-day decisions in all areas of implementation. The fact that some of them have little experience in developing countries reduces the benefits of adequately planning for a medium-term strategy. Because of their over-reliance on expatriate staff, projects have suffered from a high level of staff turnover. As a result, it is difficult to develop in-depth knowledge of the local context and to pass it on to new staff.' (ECHO, December 2001c:15)

'There is a shortage of qualified staff in both republics [of the North Caucasus]. NGOs in order to provide medical assistance are recruiting staff at a much higher salary than the one offered by the MoH. In fact, humanitarian assistance contributes to the drain of human resources from critical facilities.' (WHO, December 2002:18)

- a lack of transparency concerning decision-making (e.g. WFP, September 2002; Oxfam, July 2002; ECHO, October 2001c);
- lack of communication, and friction, between staff in the national capital and those based in the field (e.g. ECHO, December 2001b). Clearly this does not support the trust that is necessary for effective monitoring (see Chapter 2);
- several reports note that international staff spent insufficient time in the field (e.g. WFP, April and September 2002; ECHO, December 2001b).

**Despite the scale of the staffing and management problem, a rigorous analysis of why there is high staff turnover and detailed and feasible recommendations as to how this could be overcome was not in general included.** Lack of in-country security combined with very difficult working conditions appears to be one reason for high staff turnover (e.g. WFP, September 2002). The evaluation of MSF's intervention in El Salvador reports (November 2001:3): 'The stress and overwork of field staff during large emergencies is a recognised phenomenon, but little is done about this.' Several recommendations can be gleaned from the reports as to how to resolve staffing and management issues:<sup>16</sup>

- **Regional staff could be brought in.** The MSF (November 2001) intervention in El Salvador is accorded with success partly because of the use of regional staff. Regional staff are familiar with the context, can speak the language, are available within a very short time period, and are experienced and enthusiastic. Additional national staff were also recruited and made up the bulk of the field teams. In general, human resources were well managed and the regional emergency approach was seen as a significant factor in increasing effectiveness.
- **Staff could be given longer contracts.** This is recommended by UNMAS (February 2002) for consultants with mine action expertise, and by WHO (December 2002:23) which comments: 'Temporary contracts are offered and renewed repeatedly for periods of 3 to 6 months. The temporary nature of the contract deprives staff working for several years under particularly difficult circumstances from normal fixed-term staff benefits and relative job security. Most critically, it limits the ability of WHO to maintain the most qualified staff'.
- **More national staff could be hired.** For example, the WFP Angola evaluation recommends (April 2002:25): 'The CO [Country Office] should

seek to recruit a larger cadre of qualified national staff so as to reduce the high staff turnover and ensure more continuity of staff and expertise.' The evaluation of the CARE (September 2002) programme in Afghanistan also found that national staff were able to visit CARE projects which international staff could not go to for security reasons. On the other hand the evaluation of ECHO's intervention in Sierra Leone (ECHO December 2001c) found that there was a tendency to employ large numbers of national staff but that their role and key functions were not always clear and organograms, when available, were far too complex, suggesting mixed lines of authority. Comments about the difficulties of finding qualified national staff by the DEC evaluation were reported above, and there is often competition among humanitarian actors for staff.

- **Monitoring and reporting could be made more efficient, and bureaucratic demands from HQ could be decreased.** Norway Ministry of Foreign Affairs (November 2001:33) recommends that: 'This frequent changeover of staff ... could be counterbalanced with more efficient and stringent report writing, clearly following through projects so that important changes and events do not go unnoticed by later participants.' UNHCR (May 2002:39) comments: '[F]ield staff consistently complain of a constant barrage of reporting requests from headquarters, all of which are "an urgent priority" and many of which are duplicative and uncoordinated.' The recent decentralisation of several agencies has led to an uneasy relationship between HQ and the field, with HQ complaining of lack of capacity in field/country offices, and country offices complaining about the multiple demands from HQ (see also WHO, December 2002; CARE, September 2002; World Vision, June 2001; WFP, September 2002c).
- **Staff training could be increased.** Surprisingly few reports recommend this or discuss it in any detail. An exception is UNHCR's (May 2002) evaluation of its work on the protection of children. This recommends induction training for national staff and the incorporation of protection issues into generic training programmes (see also CARE, September 2002, which recommends training for national staff).

Poor management of human resources is linked to lack of connectedness/sustainability. Failure to hire and train national staff leads to a general lack of capacity building and is part of the wider lack of capacity building of national institutions –

whether at the primary stakeholder or national government and NGO level, both further discussed in the next sections. **In the light of these findings, most agencies need to re-examine their hiring policies to ensure that staff are hired and supported appropriately. The evidence from the evaluation set is that the humanitarian sector substantially underinvests in staff, who are its main resource.**

## Partnerships and Capacity Building

### 3.4.2

Minear (2002:55–6) notes on capacity building:

The humanitarian enterprise has proven itself better at delivering life-saving assistance than at strengthening local capacity. In the heat of each new crisis, the scramble to save lives often eclipses the sought-after partnership with local institutions, which in comparison seem diversionary and dilatory ... The heavy externality of international relief interventions works against the expressed preference for greater mutuality and local participation ... The fact that only modest improvement has occurred during these years casts doubt upon the sincerity of the humanitarian enterprise in strengthening local capacity and calls into question the lesson-learning process. The absence of significant improvement also points to the deeply rooted and dysfunctional power relationships that underpin the humanitarian apparatus.

Fewer reports than for *Annual Review 2002* include details on capacity building, beyond issues of staffing discussed in the previous section. One detailed discussion, however, was found in the evaluation of the extent to which UNHCR met the protection needs of children, which unfortunately conforms to Minear's pattern (UNHCR, May 2002:53):

The approach of the Office [UNHCR] has been more akin to contractors rather than partners. Some 40–45% of UNHCR's budget is channelled through implementing partners... Although there are selection criteria for NGO partners and a major partnership programme, relations tend to become complacent. Many observed that partners tend to be engaged 'on the spot' ... Part of the explanation for these trends is that there is a tendency to share expectations on administrative and budget



procedures with partners than to discuss strategic direction, policies and lessons learned. Interviews with some NGOs reported that the High Commissioner has acknowledged a tone of arrogance by the Office in relationships with civil society partners.

The evaluation of the DFID-IFRC partnership (DFID, January 2002) also reveals only partial success in strengthening IFRC National Societies. Although National Society strategic planning was found to have improved since 1999, financial sustainability has proven difficult (DFID, January 2002:18): ‘In the regions visited [Southern Africa and Central America], the financial viability of many NS [National Societies] is a major preoccupation and priority. Salaries offered are often uncompetitive and staff turnover can be high. Federation supported resource development work appears to have had relatively little success so far.’ The IFRC synthesis report (May 2002) on seasonal flooding also found that restrictive time limits set by donors can have negative consequences for capacity development for National Societies, in this case because National Societies may not be able to absorb the funds and may be pushed into unsustainable processes, although the report does not elaborate on what these processes are. The evaluation of the UN coordinated mine action in Kosovo (UNMAS, February 2002) also points out that the political decision to hand over responsibility for future mine action programmes to the Kosovo Protection Corps undermined the considerable attempts to develop local civil society capacity.

Despite this overall conclusion on the general lack of capacity building, there are examples of good practice in the reports on support of local institutions. The UNHCR (May 2002) report notes, for example, the success of the Japanese Emergency Network in Serbia. This network comprises teams of local social workers which link private and governmental agencies and services. These teams visit refugee and displaced families in their homes on a regular basis, and the network was highly commended for its work with cases of at-risk women and children. The report also notes that teams of refugee social workers were found to be central to child protection work in UNHCR’s programme.

Other good practice examples are noted in WHO (December 2002), CARE (September 2002) and the DEC (December 2001). Unfortunately these reports do not expand beyond noting that good partnerships between international and national NGOs were developed. For instance, the evaluation of WHO’s intervention in North Caucasus reports some success in capacity building in the

Ministry of Health in Ingushetia, in its TB programme. One of the central reasons for success is the support of the central Russian authorities. Outside of this there is little information on support provided to governments in terms of their capacity to prepare for and respond to emergencies, beyond that mentioned in Section 3.2.3 on health. This is either not happening or is not being captured in evaluation reports.

**Section 3.2 highlighted the lack of sustainability of sectoral interventions and the inability of relief agencies to establish connectedness. There is a parallel finding here in terms of a failure to build and support adequately the capacity of national government institutions,** even in situations where significant funding is devoted to rehabilitation activities. Even worse, government and civil society institutions may be undermined by relief interventions (Macrae, 2001) – either for political reasons, or because of lack of trust of government institutions. Without adequate government and civil society institutions, sustainable solutions to structural societal problems are unlikely to be found.

## Coping Strategies

### 3.4.3

A key area for capacity building is understanding and supporting primary stakeholder coping strategies. As pointed out by Grunewald (2001), **humanitarian action in protracted emergencies should be as much about supporting coping strategies as about feeding the hungry.** Coping strategies are linked to key areas of humanitarian action: the need for an adequate needs assessment; the ability to develop sustainable strategies that support rather than undermine what people already do themselves; and the need to treat primary stakeholders with respect. But as pointed out in the previous two *Annual Reviews*, assessment of interventions' attention to coping strategies has been at best partial.

This year a somewhat more sustained attention to coping strategies was found in eight reports, including five WFP reports. **The consensus from these reports is that coping strategies and vulnerability need to be better understood and may offer opportunities for agencies to move towards recovery activities.** As the WFP Ethiopia evaluation points out (December 2001:27): 'WFP staff do not systematically and consistently follow up during the post-distribution phase to better understand the role of WFP food aid within the broader coping strategies and

livelihood systems of the refugees. This failure hinders the development of appropriate indicators for the shift towards more recovery-oriented activities.'

Detailed attention to coping strategies was found in the opinion survey carried out with over 2,000 people for the DEC evaluation after the Gujarat earthquake, the findings of which are summarised as follows (December 2001:16): 'People constantly emphasized the need to restore livelihoods rather than receive relief and expressed some frustration that outsiders did not listen to them on this point. They wanted to receive cloth and make their own clothes rather than receive clothing, but no-one took any notice. They particularly valued cash interventions because they increased people's capacity to choose their own priorities and focus on livelihoods. Similarly distribution of building materials was seen as preferable to construction because it gave greater choice.' A similar desire for livelihood support, as opposed to free food distribution, was found in the case of IDPs in Afghanistan, where the World Vision (September 2002) intervention was assessed as being too distant from primary stakeholders' livelihood approaches.

Furthermore, the WFP Uganda evaluation (December 2001a) found that lack of attention to seasonality and harvest periods in food distribution undermined local coping strategies; the report comments on the need to understand social dynamics and build on existing distribution systems that favour the very poor. The issue of coping strategies is raised in a further three reports (Handicap International, July 2002; WFP, April 2002; WFP, September 2002), but could have usefully been further tied into evaluation analysis and findings, including more detailed recommendations related to needs assessment and future programming.

### Consultation with and Participation by Primary Stakeholders

#### 3.4.4

**In terms of consultation with and participation by primary stakeholders, this year's reports echo the disturbing story of the past two years: the limited ability of agencies to promote participation beyond implementation activities.** Consultation and participation of primary stakeholders should be a central part of needs assessment, to the extent possible, and continue through planning/design, implementation/management and evaluation. However, even in non-acute crisis situations, and despite the existence of agency policy, there are clearly major constraints to promoting more thorough primary stakeholder involvement.

**Box 3.8** ALNAP Global Study<sup>17</sup>

The concept of participation adopted by the ALNAP Global Study envisages a gradation of engagement, ranging from a minimal informing of affected populations on activities that will impact on them, through to what could be seen as a role reversal, whereby local initiatives developed by members of the affected population invite the 'participation' of humanitarian agencies while retaining control of resources and decision making. The study identifies multiple factors that influence both the feasibility and the nature of the participation engaged in, which it locates under three main headings: factors pertaining to the context (e.g. geographical, political, nature and impact of crisis); factors pertaining to the affected population (e.g. social, cultural, capacities); and factors pertaining to the aid agency (e.g. mandate, organisational culture, human resource skills, contextual understanding, existing presence). The third of these groupings is clearly the one most susceptible to change from within the humanitarian sector, where evidence from the Global Study points to a continued 'blindness' to local capacities and initiatives, and the predominance of blueprint approaches that undermine consideration of affected populations as social actors with their own insights, competencies, energies and ideas.

Although the study acknowledges the difficulties of adopting participatory approaches in certain crisis contexts, it reveals clear operational incentives in terms of increased programme/project relevance, effectiveness and sustainability, resulting from enhanced local capacities and ownership. As well such operational considerations, participation provides a means through which the humanitarian sector can demonstrate respect for those it seeks to assist, engendering greater transparency and accountability

First, two success stories. The WorldVision nutritional support programme in Kenya (June 2001) was noted as achieving good results because of the participation of mothers in the distribution of the food ration. In addition, CARE's intervention in Afghanistan was seen to have successfully supported the capacity of local institutions (September 2002:21): 'One very positive aspect of CARE's work that the evaluation team saw was the support for the creation of Community Shura. The Shura are

councils of all the adult males in the village and they elect one person to represent the village at the commune level. This type of local democracy can help to ensure that the assistance provided by agencies is appropriate to the needs of the community.' The report does not, however, go on to discuss the gender equality implications of this initiative.

**Overall, however, the reports found that needs assessments were not carried out adequately and participation in design was limited.** The report on the DEC agencies in post-earthquake Gujarat illustrates some of the reasons adequate consultation did not take place (December 2001:16, Vol 3, Report 4:1.6): 'People felt that even after they were consulted their views were not incorporated into plans, and blamed this on agencies having designed their programmes in advance ... [I]n general DEC agencies felt they were working under a severe time pressure for at least 3 reasons: First, at least in Kutch District they feared that if they did not show results quickly then other agencies would take their place. Secondly they felt a pressure to finalise temporary shelter before the monsoon, and finally most but not all DEC members were striving to maintain disbursement rates so that they could stay within the DEC 9-month expenditure period.'

**A further problem relates to the hierarchical mode of operation of some agencies.** For example, the WFP Somalia evaluation (January 2002:21) comments: 'Contrary to the principle of a "bottom-up" beneficiary-driven approach espoused by WFP, there has been a strong tendency towards a top-down approach to project planning and organization, with little community say in the procedures ... The problem of instituting a participative approach to project development and execution is not simple to solve, given the number and short-term nature of most projects and the necessary limit on field personnel.' A further issue may be trust, as primary stakeholders' opinions may not be as valued as those of other key stakeholders such as agency staff.

Similar findings about lack of adequate participation in the design phase are noted in World Vision (June 2001 and December 2001), Oxfam (June 2002, March 2002), and WFP (April 2002). Unfortunately the one report that did find evidence of consultation (Oxfam, July 2002:15) does not expand beyond the following: 'It appears that the communities were consulted in various aspects of the planning process and some responsibilities were given to the communities such as choosing who gets a latrine in the resettlement villages and selecting who gets to be a Blue Flag Volunteer.'<sup>18</sup>

## Local versus International Procurement

### 3.4.5

The issue of sources of procurement has already been raised in relation to food assistance. It is worth noting here, however, that with OECD-DAC<sup>19</sup> agreement concerning untying some elements of emergency relief – although not food aid – there may now be greater flexibility for donors in terms of procurement (CIDA, 2002).

#### Box 3.9 Excesses of International Procurement

‘Oxfam pumps, pipes and buckets flown out immediately after the disaster remained unused weeks and even months later. Water-tanks were available locally and were being distributed in large numbers by other agencies. A similar flight was sent out by Concern. Each flight costs over £100,000 and yet practically all the items could have been bought locally at a fraction of the cost, and with additional benefits to the local economy. It is hard to avoid the conclusion that these decisions were based more on a desire for publicity – or at best a culpable laziness about the use of DEC funds – rather than assessment of needs. They constitute a violation of the “humanitarian imperative”. In view of the persistence of this problem, and its potential to discredit the DEC, we suggest that any agency organising a relief flight should (for the record) submit a short justification to the DEC within a week of doing so.’ (DEC, December 2001:17)

‘The German [NGO] HELP was particularly active in the very beginning of the mine clearance campaign. Funded by UNHCR, it even helped check the office space designated for the MACC [Mine Action Coordination Centre]. Its work for UNHCR came to an end, however, in the spring of 2000, when UNHCR, in consultation with the MACC, put the continuation of the project out to tender. According to HELP, the Request for Proposals (RFP) did not specify the use of Kosovo nationals, and consequently the successful bidder came with a plan including dog teams, making 72 Kosovars trained by HELP redundant. HELP, in its final report to UNHCR, protested that there was little point in training people if they were not going to “get to work and prove their skills”. Dogs were contracted from overseas, at great expense, and “for every EDD [explosive detecting dog] contracted six deminers and their dependants have no income”. As three dog teams were brought in, 72 deminers and over 360 dependants had become destitute, according to HELP. Moreover, dogs were not as reliable as people.’ (UNMAS, February 2002:107)

At the risk of descending into journalistic excess, two of the worst cases of international procurement are set out in Box 3.9 (see also IFRC, September 2002; ECHO, October 2001a; and Oxfam, June 2002).<sup>20</sup> The need for greater local procurement, including ways in which the effects of this on the local economy can be measured, are also promoted in the context of recovery from natural disasters in the IFRC 2001 *World Disasters Report*.

Isolated examples of good practice were identified, for example the UNICEF intervention in Sudan, funded by ECHO (UNICEF, December 2001). This notes that the decision by the senior UNICEF management to allow local procurement of some materials and equipment was critical to the timely implementation of the project. Greater local procurement may be one means of supporting indigenous capacity and markets, but as the case of the chartering of aircraft makes clear (Box 3.9, and endnote 20), the pressure on agencies to maintain their profile may overcome attempts at cost-effective planning, as well as in addition undermine local capacity. Try as they might, evaluators have not been successful in shaming agencies into lower profile but more sustainable approaches.

## Gender Equality

### 3.4.6

A number of agencies involved in humanitarian action have been proactive over the last few years in terms of developing policies and tools for gender mainstreaming. For example, gender equality policies in UNDP and WFP which have recently gone to their respective boards; the FAO (2001) *Gender and Development Plan of Action*; the Inter-Agency Standing Committee's (2000) *Gender and Humanitarian Assistance Resource Kit*; and the ICRC's book *Women Facing War* (2001).

While some good practice was found this year, **the picture in terms of promoting gender equality is at best mixed and poor in many cases.** A fundamental issue – related to questions of sustainability/connectedness – is the extent to which gender equality, usually thought of as a long-term goal requiring structural social change, can be promoted through humanitarian action. A further issue is lack of prioritisation of gender equality as a key theme, indicators of which are the low level of resources allocated to gender equality, the lack of reporting on this theme, and the fact that most gender focal points are junior staff (IANWGE, 2001).

**Box 3.10 Findings on Gender Equality from WFP Reports**

Summarising the findings of the PRROs, it would be fair to say that WFP performance in relation to its Commitments to Women is significantly below expectation, for example, in relation to providing direct access for women to appropriate and adequate food, in ensuring that the ration is controlled by women, and enabling women's participation in decision-making bodies:

- 'The mission was able to confirm achievements in the field of gender in several aspects of the programme, though progress was uneven ... In many of the camps in the Great Lakes region, women are now members of food management committees, composed of three women and two men ... Additionally, in all camps, a certain percentage of rations must be distributed to women, and monitoring activities show WFP has been quite successful in this regard. Less clear is whether women have control over that food if they have help getting it home.' (WFP, September 2002:37-8)
- 'Having women head beneficiary units on distribution lists facilitates their collection and control of rations. But the issue of control of the family entitlement remains an issue and this is not being monitored.' (WFP, December 2001a:43)
- 'The PRRO design document does not include an analysis of gender relations or the special needs of female refugees ... Nor does the design methodically address WFP's Commitment to Women ... [R]ations have never been distributed directly to women [which was meant to take place under the Commitments to Women] and in practice monthly distributions are either centrally organized, conducted by male store staff who dole out each household's ration, or decentralized to blocks and conducted by male leaders.' (WFP, September 2002b:37-8)
- 'In practice, complying with WFP's Commitments to Women has not yet achieved its fullest potential and the low ratings awarded in the annex reflect the urgent need to tackle gender concerns of all types of activities and all levels of programming.' (WFP, April 2002:34):
- 'WFP's Commitments to Women are insufficiently considered in the design of the PRRO. The objectives state only that women are considered a vulnerable group and should participate in food management and distribution. No further reflections, objectives, activities and indicators to meet commitments to women are included in the design of the PRRO.' (WFP, December 2001:18, 33)



Despite this, there is more information this year than previously on the results of initiatives to promote gender equality. A detailed assessment can be found across the WFP PRRO reports, all of which pay significant attention to gender equality. The introduction of an annex organised according to the five WFP *Commitments to Women* in the WFP reports is an important innovation; this attention in itself suggests the seriousness with which WFP currently views its commitments to women. However, as Box 3.10 illustrates, WFP is some way from operationalising its commitments.

The WFP thematic review (September 2002a), based on five country studies and two desk reviews, found that, overall, *Commitments to Women* had a positive impact on corporate culture and the type and scope of gender-sensitive policies formulated by WFP between 1996 and 2001. However, weaknesses in their monitoring system, and in particular post-distribution monitoring, make conclusions about impact difficult and the findings reported in Box 3.10 make it clear that impact has at best been mixed. The thematic review has led directly to the development of the WFP *Gender Policy (2003–2007)* and *Enhanced Commitments to Women*, which include a focus on the nutritional requirements of expectant and nursing mothers and adolescent girls, as well as various measures to ensure that women play a greater role in decision making in food distribution committees and other programme-related bodies.

Assessment of contributions to gender equality can also be found in the assessment of UNHCR's implementation of policy and guidelines for the protection of refugee women (UNHCR, May 2002a), which was based on a five-country study. As in WFP, these planning documents were seen to have raised awareness among agency staff of gender equality issues; in this case implementation of the guidelines was found to be uneven and incomplete, occurring on an ad hoc basis in individual sites rather than in a systematic fashion. In addition, UNHCR's actions were found to be insufficient for provision of refugee women with equitable protection, for example against sexual and gender-based violence. Constraints identified were lack of female staff, weak monitoring, lack of a policy on gender equality, and lack of resources devoted to gender mainstreaming.

Lastly, the evaluation of CARE's programme in Afghanistan (September 2002:47) noted serious constraints in terms of promoting gender equality: 'The evaluation team found little evidence of gender analysis. Although CARE has paid particular attention to the needs of widows, the differential impact of the CARE program on

women and men, or on different age groups, appears to have received little analysis in some projects. The evaluation team found no evidence that gender was considered in CARE's emergency assessments. Although the small survey that the

### **Box 3.11** Achieving Good Coordination

#### **UN Mine Action Service (UNMAS)**

UNMAS, located in the Department of Peacekeeping Operations in the UN Secretariat, had primary responsibility for monitoring developments in Kosovo and planning the UN's joint response in mine action. There were a number of key elements in the success of the Kosovo Mine Action Coordination Centre (MACC) formed by UNMAS.

In 1998, UNMAS conducted extensive consultations which led to a detailed agreement with its partners in the UN system as well as a policy document setting out the distribution of responsibilities of each in responding to the immediate and long-term problems posed by landmines, based on comparative advantage. Agencies involved were the UN Department of Disarmament Affairs, OCHA, UNHCR, UNICEF, UNDP, UNOPS (the principal service provider for mine action and capacity building programmes), FAO, the World Bank, and WHO. ICRC and NGOs also contributed to the policy, and Kosovo was the first time it was tested on the ground.

That this policy on distribution of responsibilities was largely followed was one of the key design elements that made the mine action programme in Kosovo a success: 'It demonstrated that a partnership model, based on harnessing the competencies of various players (of which many were outside of the UN system), could lead to results that would have been hard to generate in a "command and control" structure.' (UNMAS, February 2002:8)

The success of the MACC in building a coordination framework can be attributed to the following:

- Nearly all MACC staff had credibility and competence, and the programme manager demonstrated consistent vision and leadership to both MACC staff and the wider mine action community.
- The UN had the authority normally reserved for government.

CARE Afghanistan team conducted for the evaluation team showed that CARE was meeting the needs of women as well as those of men, this did not appear to be a deliberate feature of project design.’

### Box 3.11 Continued

- The UN had some 10 years of experience in mine action, and had learned from this experience. For example, UNOPS (which was implementing most other programmes) had developed standard operating procedures for a Mine Action Centre as well as standard personnel and logistics’ needs.
- NGOs welcomed a coordinating role for MACC, and three key organisations pledged support by letter.
- Several bilateral programmes instructed their operators to accept MACC’s leadership.

#### FAO (September 2002:30)

‘In several of the countries visited, an ECU [Emergency Coordination Unit] was set up. In Rwanda, Burundi, Sierra Leone and, to a lesser extent, Eritrea, the establishment of an ECU and/or the fielding of an Emergency Coordinator was an important prerequisite for FAO to play a key role in the overall humanitarian assistance to the country. The following were the major achievements:

- Effective coordination was achieved in Rwanda, Burundi and Sierra Leone through the establishment of Coordination Committees, which appeared to be a unique and transparent forum in which the multiple actors exchanged information and experience.
- These committees have proved to be an effective management tool in fostering the harmonization of intervention strategies and operational approaches. In Burundi, management coordination became even more effective with the establishment of a joint input “pool”. This led to better coherence among interventions and avoided duplication.
- Collaboration was reflected in several joint initiatives, again in Rwanda and Burundi, in particular with WFP (e.g., joint distribution of food rations and seeds, joint needs assessment).’

## Coordination

### 3.4.7

Much has been said about poor coordination in the humanitarian sector, and this has been supported by the two previous *Annual Reviews*. Coordination failures are also noted in several reports from this year's set (e.g. DEC, December 2001; ECHO, December 2001c; ECHO, December 2001h; World Vision, June 2001; WFP, September 2002b; UNHCR, May 2002). There are also two reports which point to good practice, including how this was achieved. These are highlighted in Box 3.11 and demonstrate that, given the right circumstances, good coordination can be achieved.

## Rights-Based Approaches, Protection and Advocacy

### 3.4.8

Most reports steered well clear of rights-based issues. In fact, 53 of the 55 reports reviewed this year either did not discuss protection or, if they did, addressed it in a piecemeal fashion. However, two reports stood out in their attention to rights-based issues.

The evaluation of the extent to which UNHCR met the rights and protection needs of refugee children (May 2002) found that UNHCR has taken a number of important steps toward this goal. This included the development of a policy and guidelines on refugee children, the establishment of specialist postings at HQ and regionally, strategic partnerships for community services and education, and training and capacity building programmes such as the 'Action for the Rights of the Child' training initiative and the 'Separated Children in Europe Programme'. Despite this, the evaluation found that although children make up half of UNHCR's population of concern they are often overlooked and not considered central to core protection and assistance work. Four main factors were responsible for lack of implementation of protection activities: limited accountability; funding cuts; difficulties with mainstreaming child protection in UNHCR; and gaps in staff understanding.

Concerning accountability, the evaluation found limited compliance with UNHCR policy and a failure to follow up on past evaluation recommendations concerning the need to pay greater attention to the protection needs of children. This is partly a result of UNHCR having multiple priorities which make it difficult for staff to decide on which to act. Also, in terms of mainstreaming, the evaluation found limited integration of programmes for the protection needs of children into regular

activities. It also found confusion among UNHCR staff about what child protection means or what the policy priority on refugee children entailed.

The evaluation found that for the protection needs of refugee children to be effectively operationalised the following would be required (ibid:vii):

- the leadership and support of senior management;
- the degree to which protection staff include social as well as legal and physical aspects of protection, and integrate their work with community services and education;
- the degree to which community services staff work with community-based social systems and networks;
- the existence of strategic partnerships, especially collaboration with UNICEF and key NGOs.

The other significant report was that which evaluated DEC agencies who responded to the Gujarat earthquake (DEC, December 2001). This report uses contextual factors to analyse what is perceived as a failure of advocacy by most DEC-funded agencies. It points out that because of the political importance of Gujarat to the ruling federal party in India, sufficient relief and recovery input was likely to be provided after the earthquake by the central and state governments. NGOs were thus minor players (as already noted in the housing section, of the one million houses that needed to be rebuilt or reconstructed NGOs are slated to reconstruct only about five per cent and the DEC agencies were involved in perhaps 20,000 units).

In this context, the evaluation points out that DEC-supported agencies should have been more proactive in their advocacy work to ensure that the Indian governments' response was pro-poor. It is worth quoting this analysis at length given that it is an area missed by most EHA (ibid:28):

The main criticism of the DEC response in Gujarat was that it was not political enough, in the sense of engaging with the wider issues of the response, notably the tendency for government aid to benefit the better-off ... The ActionAid/Christian Aid partner Unnati is an outstanding

example of what could be done. By focusing on the neglected urban areas and working closely with people in informal settlements around Bhachau, they were able to bring about major changes in the process of reconstruction. Settlements which might otherwise have been wiped off the map were included in the plans and the people participated in decisions about the future.

In relation to the housing programme the report goes on (ibid:34): 'Instead of building their response on local capacities – notably the opportunities for employment and rights of poor people to compensation from government – agencies have focused on providing external solutions. By putting so much effort into these mistaken schemes they have missed the opportunity to focus on what was really needed: support to the process of self-help reconstruction – not just by providing materials but by securing entitlements.'

Similar points could be made concerning the Oxfam and Tearfund housing programmes in Bangladesh, Colombia and Serbia (see housing section above), where the agencies involved may have been better off organising primary stakeholders to gain access to basic needs. Humanitarian actors appear to consider themselves as mainly service providers, and there is a subsequent loss of opportunity to support the organisational capacities of primary stakeholders.

### Impact on the Environment

#### 3.4.9

Although a central cross-cutting theme in development, the environment is usually paid little or no attention in EHA, perhaps because it is considered that humanitarian action will have limited impact on the environment or because data is not available. However, this year there is some attention in the WFP PRRO reports, as follows:

- The Great Lakes report (WFP, September 2002) accords the intervention in the region with success due to UNHCR and WFP policies for environmental protection and reclamation. In order to reduce fuel consumption, WFP provided a food basket that demands less energy for cooking, including milled grain, and women were trained in fuel-efficient cooking. WFP was also involved in reforestation, reclamation and soil conservation, and the

evaluation found that the extent of environmental degradation through deforestation in camp areas was significantly lower than in and around many other African refugee sites.

- Energy saving stoves have been successful in Angolan, Ugandan and Tanzanian refugee camps (WFP, December 2001a, April 2002 and September 2002), but not in Ethiopia (WFP, December 2001) – although it is not clear why this difference exists, except that there may have been some variation in design.

The evaluation of CAFOD's (July 2001) intervention after the Mozambique 2000 floods has a specific focus on the environment. This report found that no environmental impact assessment was carried out. Furthermore, the attempt to meet shorter term needs, such as emergency shelter provision, was likely to have negative effects on the environment because of the widespread cutting of poles that could lead to worse flooding and siltation of the Limpopo river in the future.<sup>21</sup> This ties into the discussion earlier in this chapter on local procurement, and the CAFOD report notes the need to ensure that local procurement supports environmental sustainability.

### Results-based Planning<sup>22</sup>

### 3.4.10

**Reports were so consistent in their criticism of agency monitoring and evaluation practices that a standard sentence could almost be inserted into all reports along the lines of: 'It was not possible to assess the impact of the intervention because of the lack of adequate indicators, clear objectives, baseline data and monitoring.'**<sup>23</sup>

The ad hoc nature of monitoring comes across in several reports, for example the evaluation of the Norwegian Red Cross (November 2001) which comments on the lack of a format for either situation reports or end-of-mission reports, the informal nature of reporting, and the resulting loss of knowledge. Quotes concerning poor monitoring are highlighted in Box 3.11 and reinforce the main findings from Chapter 2 concerning constraints to effective monitoring. The one report that notes high quality monitoring (CARITAS, January 2002:25) unfortunately does not expand on the comment that 'the quality of the monitoring reports tends to be high', in terms of how this unusual achievement was accomplished.

**Box 3.12** Poor Quality of Monitoring Across the Humanitarian System**Lack of monitoring**

'There is no baseline data, no data on food security is being collected and there is no monitoring of the impact of food assistance or the exclusion of certain refugees from assistance. There is no monitoring of potentially marginalized groups.' (WFP, September 2002b:20)

'Knowledge of the degree of self-sufficiency among the target refugee population is limited and the mission could not find indicators that the nutritional requirements of the refugees have even been analysed in a systematic way following UNHCR/WFP guidelines.' (WFP, December 2001:9)

'It is impossible to measure the Public Health impact of the programme, as appropriate data has not been collected.' (Oxfam, June 2002:10)

**The wrong kind of monitoring**

'In the absence of a logical framework analysis the current monitoring system functions in a vacuum. Monitoring activities focus on input and output monitoring. Little monitoring is undertaken beyond food distribution monitoring ... WFP progress reports are mostly quantitative in nature and provide little information on qualitative achievements.' (WFP, April 2002:20)

**Too many demands on field staff**

'Many field staff are already overwhelmed with the monitoring requirements associated with day-to-day food movements, arrivals and distributions. They do not easily see how to shift their focus towards monitoring the more qualitative aspects of the PRRO.' (WFP, September 2002:19)

**Lack of use of data**

'Although the recommendations contained in many of the [Joint Food Assessment Missions] reviewed by the mission were found to be quite useful and concrete, there was little evidence of any systematic follow up of tracking of recommendations on the part of management. Part of the problem may relate to the fact that many of the reports fail to clearly identify which party is responsible for following up on actionable items and within what timeframe.' (WFP, September 2002:23)



Alongside the general weakness in results-based planning, some additional areas could be discerned from this year's evaluation set as follows:

- **Agency staff are already overloaded and do not tend to or do not have the time to prioritise monitoring.** For example (ECHO, December 2001c:12): 'Collecting data on effectiveness has been problematic. The heavy demands upon field staff and organisational structures preclude adequate and reliable data collection.' And WFP (January 2002:19): '[R]eporting formats ... have been found to be excessively lengthy and to use indicators which are not always seen to be relevant by WFP field staff.'
- Monitoring of quantitative elements of interventions such as the amount of food distributed is in many cases adequate; **qualitative elements such as who actually benefits from the food seem to be beyond the capacity of most agencies.** This is noted in several WFP reports and also, for example, ECHO (December 2001b:20): 'Although the number of children who have been reintegrated with their families/foster families/the community is reported, none of the projects has done real analysis on the quality of the reintegration that would help adapt the strategy.'
- **In some cases too much or inappropriate data is being collected, or the data that has been collected is not being used** (e.g. Oxfam, July 2002), partly because of a lack of capacity to carry out analysis.
- **Agency staff often don't have the skills to carry out adequate monitoring, nor have agencies in general provided adequate resources for training.** As the report on World Vision's programme in Ethiopia (December 2001:9) notes: 'It was the first time staff had implemented a targeted food program that involved monthly screening of participants ... Data had to be collected and records reviewed to determine program defaulters, deaths and discharges. Unfortunately the staff training, delineation of program criteria and guidelines, as well as the monitoring of program implementation, ongoing activities and monthly records was inadequate.'

As found in *Annual Review 2002*, a large number of reports recommend the introduction of the log frame approach to help overcome general weaknesses in

results-based planning (e.g. WFP, January 2002, December 2001a, April 2002; Oxfam, July 2002). It is not clear, however, how use of log frames will support more attention to qualitative areas, which has been identified as one of the central gaps in results-based planning. In addition, qualitative questions related to social process and causality can only be captured by skilled observers and are most effectively captured through ongoing communication with primary stakeholders – including where possible primary stakeholders as participants in results-based planning. Unless these key social science skills are strengthened in agencies and their partners we will remain partly in the dark about key aspects of humanitarian action – such as why certain groups benefit and how lessons can be learnt.

There is also limited information on attempts made to strengthen national institutions that could carry out results-based functions, or on any attempts to include primary stakeholders in monitoring and evaluation (part of the wider inability to develop capacity). For example, the evaluation of World Vision's intervention in Ethiopia comments (December 2001:2): 'World Vision Ethiopia (WVE) monitors the food security situation in its twenty-three operational areas ([Area Development Programmes] ADPs) on a semi-annual basis. This surveillance includes the collection of data on crop prices, agricultural conditions and the nutritional status of children under five. WVE is the largest contributor of data to the government's early warning system.' The fact that an NGO is the largest contributor to the Ethiopian government's database suggests that capacity remains with an external agency rather than the government.

## Synthesis Summary and Conclusions

3.5

Key points from this year's review are that:

### Sectoral findings

- Humanitarian action continues to be successful at saving lives, reducing or stabilising morbidity/mortality, and providing basic needs related to water/sanitation and health. There is much good practice on which subsequent interventions can be built.

- **Food aid interventions** generally meet their primary objective of feeding the hungry, although there are serious questions concerning whether targeting is appropriate, and the lack of data makes drawing conclusions difficult. The questions of intra-community and intra-household distribution are areas about which little is known, despite being recognised as key by several evaluations. Local procurement and alternatives to food aid such as cash-for-work are not being adequately considered in interventions.
- **Water and sanitation interventions** were also generally successful in meeting physical targets, but sustainability of facilities is a major problem. Findings on participatory management mechanisms for water facilities in particular point to poor results as well as lack of institutional learning. This may result partly from a lack of guidance in this area.
- The objectives of **health interventions** were mainly stated in terms of stabilising mortality and morbidity and ensuring that infectious diseases were kept under control. These objectives were mainly met, but there was a failure to build local capacity; what's more, there is a lack of guidance as to how this can be achieved.
- **Housing** is the most problematic sector, and the results of reports on housing support the findings of *Annual Review 2002* concerning the inability of relief interventions to facilitate sound housing reconstruction.

### Cross-cutting findings

- A number of institutional barriers exist to more effective relief, in particular **personnel/management issues** (communication problems between HQ and the field; high staff turnover; short-term contracts; too many expatriate staff), and **poor results-based planning and monitoring**. This notwithstanding the dedication of often over-worked and over-stressed staff. Despite the scale of the staffing and management problem a rigorous analysis of why there is high staff turnover and other problems, and detailed and feasible recommendations as to how this could be overcome, was not in general included.
- Humanitarian action focuses on traditional mechanisms of aid such as providing food aid or building houses. **There is no evidence of a move to**

**more rights-based approaches.** Fifty-three of 55 reports either did not discuss protection or covered this issue in a piecemeal fashion.

- **Agencies are mostly non-participatory** and have not included primary stakeholders, in particular in design and planning. **Consequently indigenous coping strategies are insufficiently understood** and may be undermined by relief interventions.
- **Greater local procurement** may be one means of supporting indigenous capacity and markets, but the pressure on agencies to maintain their profile may overcome attempts at cost-effective planning and, in addition, undermine local capacity.
- **Successful initiatives that promote gender equality are few and far between.**
- **Agency staff either cannot prioritise monitoring or do not have the skills to collect and analyse data in a way that is useful, particularly in qualitative areas related to social process.**

Evaluation reports paid substantial attention to **LRRD and sustainability/connectedness**, and concluded that, in spite of the fact that much relief funding is used for rehabilitation and recovery activities, agencies have not in general been able to promote connectedness/sustainability in any of the main humanitarian sectors – a finding in the detailed reviews by Minear (2002) and Macrae (2001). The main reasons given for this in the reports are:

- Limited attention is paid to supporting and building the capacity of government and civil society.
- Indigenous coping strategies of primary stakeholders are not well understood.
- There is inadequate local procurement.
- Staffing profiles emphasise the hiring of international staff even where it may be more appropriate to hire local staff.

- Donors establish restrictive time limits on expenditure. This means that agencies either cannot absorb the funds or use them to maximum effect (for example, DEC, December 2001; IFRC, September 2002).

Systemic problems, such as the lack of operation and maintenance of water structures or information on when to move from the relief to the rehabilitation phase, suggest that **there is a major lack of guidance and training on how to link relief and rehabilitation – i.e. on how to operationalise LRRD policies**. All humanitarian agencies face persistent problems with this issue. This suggests that it should be a central focus for training and capacity building within the humanitarian sector.

In relation to longer term strategic planning, the evaluation of the CAP argues that the CAP should be more than a shopping list for donors and should be the document that reflects strategic planning on LRRD (OCHA, April 2002:38): ‘A well-done CAP should be able to articulate how relief, rehabilitation and development activities inter-relate and work towards shared strategic objectives.’ This makes intuitive sense.

Failure to promote sustainability/connectedness is tied to another theme from this year’s reports: poor understanding of local context and situation analysis. Unless agencies take local context into account they will be unable to understand the constraints and potentials of programming, including capacity building. Related to this is a lack of understanding of social process. Agencies that have been set up to deliver goods such as food are not in general effective at collecting and analysing information on social process. Without this qualitative information it is impossible to understand how interventions could work better, why they succeed or fail, and who benefits. A good example is food distribution. Reports are clear that food aid is distributed in a manner that is reasonably consistent with agency objectives; what happens after delivery in terms of who benefits from food aid remains for the most part a mystery. This is where effective monitoring should come in, but as Chapter 2 and the discussion above notes, monitoring is one of the weaker areas of agency function.

Having said this, there was greater attention to issues of causality – one area of social process – in this year’s reports in comparison with those from last year. *Annual Review 2002* argued that a general lack of attention to causality in EHA is one of the

main factors hindering a better understanding of why results are achieved, which is in turn a major constraint to lesson learning. The central reason for a greater focus this year was that WFP PRRO evaluations had taken this on board; several of these reports (for example, December 2001a and January 2002) included a section on 'Factors in the Effectiveness of the PRRO', under which was included human resources, targeting, capacity building, monitoring, security, and finance. Findings on causality from these reports are covered in relevant sections above. Other evaluation offices could usefully include such a heading in their own terms of reference, explaining that this section should bring together all of the reasons why an intervention has achieved or not achieved particular objectives, and subsequently draw lessons learned from this.

As EHA moves tentatively towards more of a lessons learning approach, partly as a result of having recommendations ignored over many years, its accountability function should not be forgotten. Rather than losing heart with the accountability approach and veering too strongly towards an as yet largely untested lessons learning approach, agencies should consider how to bring about an appropriate balance between these two evaluation functions. This means considering in more detail how evaluation results are going to be used, and then working back from this as to what mix of accountability and lesson learning is most likely to promote this use.

The fact that many issues recur, such as lack of attention to gender equality, lack of participation of primary stakeholders in design and planning, poor attention to protection, and the distorting effects of tied aid, does suggest major accountability problems in the humanitarian system. Rather than results-based planning leading to greater accountability for results, it may be establishing smokescreens that deflect attention away from the fact that accountability systems in humanitarian agencies remain weak. Much of the focus of results-based planning is on the development of tools and systems, and the production (rather than the use) of data and reports, while the staff assessment side of results-based planning is left largely untouched. There is still a long way to go before greater accountability is integrated into agency human resource systems. There is a delicate balance to be maintained here: too much focus on individual accountability may lead to a culture of blame and lack of lesson learning and experimentation, as discussed in *Annual Review 2002*.

Evaluators have a responsibility to continue to point out problems with EHA, to lobby for changes to be made and for their recommendations to be taken seriously.

Rather than being an ‘objective’ and external observer, it is increasingly understood that evaluators have a role in ensuring that change takes place, that poor practice is changed, and that good practice is rewarded. Just like agencies, evaluators need to be accountable not only to those who are funding them but also to primary stakeholders in whose name they are carrying out their work.





## Chapter 4

## Conclusions

## Overview

4.1

This final chapter draws out the synergies between the various strands of this year's Annual Review. First it reflects on the findings that concern monitoring within the context of the sectoral picture provided by the evaluation synthesis and meta-evaluation. Then it provides an agenda for improving monitoring, for both donors and operational agencies. The chapter should be read in the light of the Annual Review's aim to meet the needs of different audiences – those already knowledgeable about monitoring and evaluation, and those less so.

The constraints almost all agencies face in developing adequate monitoring systems constitutes part of a wider problem concerning the establishment of adequate management and reporting systems. The question addressed at the end of this chapter is how strengthened monitoring can support more effective humanitarian action. This is considered key for improving both humanitarian action and its evaluation. If agencies and governments cannot learn both between and during emergencies their response will continue to exhibit the generic problems identified in consecutive ALNAP Annual Reviews.

## Assessing the Performance of the Sector

4.2

This year's assessment involved 55 evaluation reports, allowing general conclusions to be drawn on the operation and achievements of the sector in 2002. What constitutes success in humanitarian action is open to varying interpretation. From the perspective of the stated objectives of many interventions – saving lives and/or maintaining nutritional levels – humanitarian action achieved its primary objectives. However, consecutive syntheses have revealed that humanitarian action is unable to create sustainable services or rebuild livelihoods in the medium to longer term. Successive Annual Reviews have given indications as to why this is the case by outlining systemic structural barriers which, if left unchecked, will continue to impede performance.

Nevertheless the achievements of humanitarian action should not be underestimated. In each of the four main sectors of humanitarian action – food aid, water and sanitation, health, and shelter – there were notable successes. Almost one half of the evaluations reviewed included a focus on food aid, and results illustrated the effectiveness of food aid in feeding the hungry poor in a number of challenging environments. However, several evaluations suggested that many vulnerable groups, particularly those outside refugee camps, may be bypassed by food aid; often the data is not available to substantiate who has actually benefited. Clearly more attention needs to be given to using data to ensure appropriate targeting. The importance of attempting greater local procurement was also raised in several reports.

Water and sanitation interventions were successful in meeting physical targets, for example, installing handpumps and protecting water sources. Health interventions largely met their short-term objectives, in particular stabilising mortality and morbidity rates and ensuring that infectious diseases were kept under control. While, as in previous years, temporary shelter and housing interventions were evaluated as significantly less successful than other sectoral initiatives, there was also some good practice that could be highlighted. An important finding is that self-construction, supported by external agencies, is likely to be the most effective reconstruction approach – as long as the fact that some vulnerable groups may not be able to reconstruct their housing is taken into account.

Problematic areas also re-emerged this year: limited attention to rights-based approaches and protection, with only two of 55 reports adequately evaluating protection; low levels of participation of primary stakeholders, in particular in planning and design; and poor quality of programming in relation to gender equality.

4

This year's evaluation set also substantiated in significant detail worrying findings from the previous Annual Reviews. Most noteworthy, and a consistent theme, is the continuing lack of attention to sustainability/connectedness and the inability of the humanitarian response to establish a basis for longer term development processes – in essence, therefore, providing a sophisticated band aid solution. This problem plays out in a number of interconnected ways, in particular:

- 1 the failure to develop and support national and local capacity;
- 2 the tendency to build physical structures, such as hospitals, while paying inadequate attention to maintenance and operation;

- 3 lack of local procurement, which could prove both more cost-effective and support local capacity;
- 4 hiring of international staff when equally qualified national staff are available.

A compounding factor is underinvestment in capacity building of frontline agency staff, including lack of guidance and training on how to link relief and rehabilitation. All humanitarian agencies face persistent problems with this issue. This suggests that it should be a central focus for training and capacity building within the sector.

Overall the emphasis in humanitarian action remains very much on external agencies entering often unfamiliar localities, providing instant relief, and then moving on. The capacities and contributions of governments and primary stakeholders – in the latter case usually far more important than external agency input – are systematically ignored. The consensus from several reports is that coping strategies and vulnerability need to be better understood and may offer opportunities for agencies to move towards recovery activities and hence bridge the LRRD gap.

### Developing and Supporting Agency Staff Capacity

#### 4.1.2

Problems related to high staff turnover are increasingly acknowledged as each Annual Review is published, and this is one of the first systematic analyses of this issue across the sector as a whole. High levels of staff turnover at both field level and HQ, pressurised work environments, excessive working hours, too many staff on short-term contracts, and not enough emphasis on employing regional and national staff are generic problems which lead to less effective performance and limited ongoing learning. As a consequence the importance of establishing sound information flows, including effective monitoring, are downplayed. A partial solution may be to hire more national staff. While there is often competition among external agencies for national staff during emergencies and rehabilitation, the knee-jerk reaction of external agencies is to send in expatriates. A longer term solution would be to build and support government capacity to respond – an area that is rarely covered in EHA.

Having said this, what is particularly striking (and paradoxical) is that the successes of humanitarian interventions appear to a large extent to be due to the enormous

efforts and commitment of humanitarian personnel rather than because of smoothly functioning aid systems. Last year, the Annual Review concluded that humanitarian staff compensated for inefficiencies and failings in the sector. The same is true this year. This is reflected in Chapter 2 where it was found that good monitoring depends as much as the quality of staff as it does on the systems themselves.

The most important area for agencies to focus on over the short term, therefore, is strengthening field staff ability to carry out participatory social science analysis – who is affected, why, what are their existing coping strategies, and how will any intervention support or hinder longer term development? Field staff also need to know how to analyse and package this information for agency consumption. Agencies should assess the current capacity of their systems in this area, and adjust accordingly.

Of course training will not be a universal panacea. No matter how strong the situation analysis there will be continuing pressure for profile, for example, to import expensive relief goods emblazoned with an agency's logo (ironically the greater focus on results, of which improved monitoring is a part, has also led to a need for greater visibility among the public, donors and executive boards). But even in this context training can support better overall quality of performance.

## Changing the Financial Planning Horizon

### 4.2.2

#### 4

High staff turnover is directly influenced by the funding arrangements of many donors and operational agencies. Although the complexities of the relief-reconstruction-development 'continuum' are increasingly understood, leading to some modification of institutional arrangements within donor organisations, a high proportion of humanitarian assistance is still planned with a short-term perspective and funded through six to 12 month project grants. Agencies have not, or have not been able, to make substantial changes to funding patterns – mainly because of political and institutional reasons – despite this having been highlighted in a number of evaluations included in the Annual Reviews.

In practical terms there is a need to extend the financial planning horizon of agencies to create more flexible arrangements so that, among other things, staff can be hired on longer contracts. This will support work in three areas which have been

found in the Annual Reviews to be relatively weak: understanding of the context of the emergency; building partnerships with national and local institutions; and engaging with the recipients of humanitarian assistance. This in turn would increase the potential for national capacity building.

## Meta-evaluation of Evaluation Quality

4.3

The meta-evaluation, included in the coloured pages at the back of this Annual Review, provides an overview of the quality of evaluations based on an assessment against the ALNAP Quality Proforma (QP). Strengths found in previous years were repeated in this year's evaluation set, in particular the ability to assess management and human resources and to measure agency performance against several of the DAC criteria. There was also a significant improvement in the assessment of sustainability/connectedness. However, previous weaknesses were also repeated and analysis of data over the three years found no significant improvement; in some cases even a decline in quality of evaluation reports. Particularly problematic areas continue to be:

- 1 the failure to substantiate the findings of reports, often because of inadequate delineation of methodology;
- 2 the failure to consult adequately with primary stakeholders or to detail consultation that has taken place in a way that informs report findings;
- 3 poorly formulated recommendations.

A number of recommendations are made in the meta-evaluation section as to how to improve evaluation practice. Most importantly, the limited substantiation of key findings with adequate evidence undermines the credibility of many reports, and in fact EHA itself. There is also some evidence to suggest that, based on feedback from country and sector specialists, evaluations are providing a somewhat rosy picture of agency performance – an area the Annual Review must consider in future issues. This relates partly to the evaluation process where internal evaluation offices commission evaluations from 'independent' consultants, who then proclaim that the

evaluations themselves are ‘independent’ thus ignoring the many levels of negotiation and debate which usually take place both within agencies and between evaluation offices and evaluators before evaluations are published. It also relates to the narrow scope of many evaluations, which focus on output-related issues such as providing food or water to the exclusion of rights-based issues such as protection and gender equality.

EHA is of course not alone in facing these methodological and procedural challenges – they are an inherent part of any evaluation process. In our case, pointing out system-wide areas of weakness appears to have had limited impact in terms of changing practice as evidenced by the consistent poor performance in some evaluation areas highlighted in the meta-evaluation. Many of the problems highlighted in the Annual Reviews may already be known to agencies, but they cannot be fixed because of lack of resources and capacity and/or intra-agency political pressures. For this reason, this year ALNAP is following up with individual agencies as to the results of the quality of their evaluations as well as initiating a discussion of minimum standards for EHA in key areas – for example, the need for substantiation of conclusions, attention to gender equality, and adequate and fully documented consultation with primary stakeholders. This will be one step toward further professionalisation of EHA along the lines of, for example, the African Evaluation Association, which has adapted the evaluation standards promoted by the American Evaluation Association for its own purposes.

## Monitoring

4.4

### Refocusing the Reporting Agenda and Learning from Monitoring

4.4.1

Chapter 2 reported on an ALNAP research project to examine the contribution that monitoring can make to more effective humanitarian action and EHA. The focus on monitoring is part of a growing trend in EHA to look outside the ‘evaluation box’ toward more innovative means of assessing results, promoting lesson learning and, as a consequence, improving performance.

The added value that monitoring can bring to the humanitarian endeavour lies in its ability to assess and reassess continually the relevance and impact of interventions. In addition, monitoring can examine social process – that is, the complex set of relations between agencies and primary beneficiaries, on which intervention results are largely dependent; evaluation as it is currently carried out as a ‘one-shot’, usually brief effort, cannot do this in any detail. Monitoring is therefore crucial for organisational learning processes. Given the current low quality of monitoring there is significant potential for improved monitoring to lead to improved performance – i.e. investing in monitoring is likely to both improve results and be cost-effective.

The short research project on which Chapter 2 is based found that, in current practice, monitoring is usually perceived as less important and more routine than evaluation. This is because monitoring has tended to focus on largely internal functions such as financial inputs and physical outputs, while evaluation has covered higher profile external areas including outcomes and impact. However, evaluation often happens too late to improve performance of ongoing humanitarian action, although it should facilitate learning for future operations. Four connected phenomena are leading to a rethinking of monitoring and evaluation functions in relation to humanitarian action:

- 1 Increasing understanding that recovery from emergencies does not normally follow a linear progression from relief to development. In the case of complex emergencies, but also to a large extent natural disasters, many poor and vulnerable households do not appear to move along any kind of continuum. The idea that emergencies have a definite end at which point an external evaluation can take place is, therefore, increasingly questioned. The varied stages of response to the emergency therefore require different reporting functions.
- 2 The recognition of the importance of understanding process and feeding ideas back into ongoing interventions on a continuous basis, leading to an uptake in the use of, for example, RTE, by larger agencies such as UNHCR and WFP. The meta-evaluation of evaluation quality found that of 36 evaluations for which data was available, 16 evaluations were conducted on ongoing interventions. Moreover the number of evaluations carried out during ongoing operations is probably due to the long-term nature of many interventions.



- 3 Increased attention to lesson learning, to which monitoring can be central, and also one reason for greater focus on RTE.
- 4 The availability of methods and techniques for ongoing assessment of social process, such as utilisation-focused evaluation, participatory rural appraisal, and process monitoring, which make it difficult to argue that sound methods for monitoring are unavailable.

## Current Monitoring Frameworks

### 4.4.2

Chapter 2 found that there are two main complementary foci for agency monitoring:

- 1 Situation, or contextual, monitoring
- 2 Results/performance monitoring

While in a holistic monitoring system both of these approaches are used, the extent of their use will depend on the type of emergency. The longer an operation, the greater the importance of moving from monitoring inputs and outputs to processes, impacts, and strategies, and the greater the importance of maintaining a realistic analysis of the causes of the problems humanitarian interventions are meant to tackle. Evidence from evaluations suggests that it remains a constant challenge for agencies to move their monitoring from administrative or logistical issues to those related to impact and strategy. A closely related challenge is to maintain an updated situation analysis.

### Situation and Performance Monitoring in Different Kinds of Emergencies

#### *Monitoring in acute crises*

Although findings from the synthesis of evaluations in Chapter 3 suggest otherwise, most of the head offices of agencies interviewed as part of the research for Chapter 2 felt that their monitoring systems worked reasonably well in acute crises, mainly because of the strong interaction between HQ and field staff and a widespread use of informal means of monitoring (especially phone calls, e-mails, and field visits). However, the discrepancy between the synthesis of evaluations and findings from the

monitoring survey may be explained by the fact that what characterises the early stages of an emergency is the very strong demand for information, and hence a strong degree of usually informal communication. While HQs feel they are receiving the information they need, this does not necessarily mean that adequate systems are in place for the transference of both formal and informal information.

#### *Monitoring in longer term emergencies*

A key feature of more chronic or longer term emergencies is the importance of a strong situation analysis as well as the ability to constantly update and refine this. From a monitoring perspective there may be little difference between longer term emergencies and development interventions. Donor fatigue is an increasingly common element in longer term emergencies: to counter this, operations need greater capacity to monitor and report on their impact and to make linkages to local civil society and government.

#### *Monitoring for protection*

The protection agenda has greatly increased the range of issues humanitarian workers are now expected to monitor. While the focus used to be on the successful delivery of assistance to affected populations the agenda has shifted to include the extent to which the rights of civilians are respected, particularly in armed conflicts.

Chapter 2 summarises the numerous types of monitoring currently taking place in the humanitarian sector, and reached three main conclusions:

- 1 accountability is for the most part upward to meet agency needs, rather than downward to primary stakeholders;
- 2 methods tend to be either quantitative or a mix of quantitative and qualitative;
- 3 there are multiple monitoring approaches, many of which overlap. This in itself is part of the problem – each individual agency has developed its own system and approach, leading to a lack of harmonisation, over-complexity, and multiple monitoring requirements from different donors.

### Current Performance and Constraints in Monitoring

#### 4.4.3

The research carried out for Chapter 2, as well as the synthesis chapter, have established that current monitoring practices reflect the problems inherent with systems in the sector as a whole: limited primary stakeholder involvement; lack of adequate administrative systems; poor levels of communication and inadequate feedback; and lack of qualified staff. Monitoring is in general considered a low-priority activity, is under-resourced, and is confined to routine and bureaucratic activities. Straight-jacketed to ensure upward accountability, monitoring fails to build meaningful and continuous information exchanges between stakeholders. It therefore currently offers limited potential for supporting ongoing learning.

The synthesis chapter, as well as previous analyses such as that related to Kosovo in Annual Review 2001, found general weakness in monitoring practice. The ad hoc nature of monitoring comes across in a number of reports this year, and the one report that notes high quality monitoring does not elucidate how this was achieved. On the other hand the analysis carried out in the meta-evaluation of evaluation quality showed that the assessment of monitoring performance in EHA is relatively strong and that this area rated well above the average for areas covered in the ALNAP QP. Nineteen per cent of evaluations were rated as 'good', and 44 per cent as 'satisfactory' in relation to their analysis of the intervention's monitoring and/or real-time evaluation mechanisms, and the effect of these on intervention results. Along with the research carried out for Chapter 2 we therefore have a sound basis for our conclusions concerning monitoring.

A particular problem that needs to be highlighted is that of staff overload. Agency staff are already heavily committed with reporting and other requirements and do not prioritise monitoring. We may now be faced with the irony that results-based management systems, which were meant to enhance accountability and decentralised management, have simply added to field staff burdens and made staff performance less rather than more efficient. The added gains in knowledge concerning impact may not balance this tendency, especially if results-based information is not being used. Earmarking of funds also adds to overloading as it requires extra levels of reporting. The current situation can be summed up as: more demands, less resources, and confusion as to the value of monitoring in the first place.

One conclusion from Chapter 2 is that many agencies are requiring staff to monitor against a wide range of commitments, including in relation to development goals.

Monitoring now needs to cover commitments to international protocols and resolutions, including minimum standards; ethics in humanitarian response; gender equality; and human rights and protection issues. International initiatives like the Red Cross/Red Crescent Code of Conduct, the Sphere standards and indicators, and the Humanitarian Accountability Project (HAP), greatly increase the range of issues to be covered. All of these have perfectly valid aims. The question is whether the agencies concerned can field sufficient human resources to ensure that they are all monitored effectively. If the human resources cannot be found and deployed, how in practice should staff cope with these different and increasing monitoring requirements?

Other problematic monitoring areas are summarised in Table 4.1.

| <b>Table 4.1 Problematic Areas in Monitoring</b> |  |
|--|--|
| <b>Planning</b>                                  | <ul style="list-style-type: none"> <li>● Lack of planning; absence of logical framework or similar planning tools.</li> <li>● Insufficient linkages between planning, implementation, monitoring and evaluation in the project cycle and management practices.</li> </ul>  |
| <b>Focus</b>                                     | <ul style="list-style-type: none"> <li>● Narrow focus on monitoring for donors.</li> <li>● Focus on inputs and outputs rather than impact.</li> <li>● Focus on quantitative analysis rather than analysis of social process.</li> </ul>  |
| <b>Resources</b>                                 | <ul style="list-style-type: none"> <li>● Financial: monitoring needs to be included in budgets.</li> <li>● Human: lack of investment in training and capacity building.</li> <li>● Systems: poor performance of monitoring systems.</li> </ul>   |
| <b>Stakeholders</b>                              | <ul style="list-style-type: none"> <li>● Lack of involvement or strengthening of capacity of government institutions and implementing partners.</li> <li>● Limited participation by primary stakeholders in the monitoring process at various stages (e.g. design, implementation, feedback).</li> <li>● Ineffective communication across different stakeholders.</li> </ul> |
| <b>Constraints linked to emergency situation</b> | <ul style="list-style-type: none"> <li>● Security and access.</li> <li>● Fluidity of the situation.</li> <li>● Lack of baseline survey/data against which to monitor.</li> </ul>   |

## How to Improve Monitoring

### 4.4.4

#### Improving Trust and Feedback

A key issue in acute crises is the degree of trust between field staff on the ground and their colleagues in country, regional and head offices. Frontline staff have to both understand and feel comfortable with their agencies' monitoring systems. However, what emerges from our analysis is a picture of monitoring as a control tool, extracting data to verify the status of specific projects. The accountability function of monitoring eclipses the learning one, and compliance overrules adaptation.

A key indicator of the strength and relevance of a monitoring system is the quality of feedback that those generating the information receive. Feedback has multiple functions. It can be limited to just basic quality control and queries about the data submitted. But it can also provide field staff with real encouragement to improve the quality of information they are supplying and to acknowledge their accomplishments. Feedback, of course, is closely linked with learning.

Related to trust is perceived use and ownership. Staff are probably more likely to spend time and effort collecting data if they are confident that those whose job it is to analyse and use that data will take account of it and act upon it if necessary. Note the vicious circle: ineffective monitoring systems are not likely to be used fully, and the fact that they are not used discourages staff from contributing to them. In fact lack of use of findings seems to tie monitoring and evaluation of humanitarian action together.

Monitoring is not simply about collecting information, but is about the capacity to circulate it swiftly to those who need and can act upon it. While formal monitoring data sits in donor reports, managers make their decisions on the basis of more informal consultations with staff or implementing partners. The challenge would then be to institutionalise and professionalise this informal communication so that it can support learning more fully.

#### Determining and Explaining Information Needs

Staff developing generic monitoring systems are not always the best communicators of the uses to be made of the information gathered. In their guides and manuals agencies need to do a better job of illustrating:

- 1 What needs to be monitored, and why?
- 2 Who should be involved?
- 3 What skills are needed to do the job?
- 4 How will the information be used?

The information needs as perceived by fieldstaff are often not prioritised. In the case of food aid interventions in Malawi, and under the umbrella of ALNAP's LSO over the last six months, field staff from NGOs were able to meet and define their own priorities for monitoring as part of a wider process of information sharing and learning facilitated by the LSO. The field staff themselves prioritised targeting, especially economic status; household size; vulnerability, malnutrition and admission to Nutrition Rehabilitation Units; asset sales; and the impact of rations received, including the use made of food supplied and the impact on local production.

### Simplifying Systems

Donors should consider simplifying reporting needs for fund recipients, in particular in relation to earmarked funds. Within individual agencies one partial solution to staff overload may be to nominate staff with appropriate skills and experience to draw together monitoring information and summarise it regularly in a way that will be useful at the organisational level.

The review of collective monitoring initiatives within the humanitarian sector included in Chapter 2 revealed mixed findings in relation to both what constitutes collective monitoring and whether this is desirable or indeed feasible. OCHA-sponsored Humanitarian Information Centres (HICs) appear to offer good potential for stronger collective monitoring, at least in terms of a more organised and collective situation analysis across the whole sector. HICs aim to manage information for systematic sharing of information, which should lead to more informed decision making.

It may, however, be over-ambitious to propose convergence in monitoring systems across the sector, but there is certainly a need for greater debate on the overall objectives of monitoring. ALNAP should consider promoting collective work for the improvement of monitoring in a similar way to its work on defining standards

for, and reporting on, the quality of EHA. However, before ALNAP decides to follow up on its preliminary study with any kind of training programme, the Network may wish to consider helping its members design an ‘auditing’ system which reviews how well an agency’s planning and monitoring system responds to different types of emergencies.

### Making Connections with Civil Society and Government

Accountability lines in most donor-funded operations tend to be vertical, with information going back up the line either to a donor agency or to a country, regional, or head office. However, if they are to develop adequate situation analyses, agencies also need to develop strong horizontal linkages with governments, civil society and other agencies. Better use could be made of national capacity, or building of that capacity could be supported where it does not exist.

### Shifting the Focus to Impact

Given the difficulties of LRRD, as highlighted yet again this year, monitoring should be able to focus on what actually happens after emergency aid delivery has taken place. Monitoring could provide contextual information – ‘information bridges’ – between these different activities.

As already noted, perhaps the most important added-value of monitoring (in relation to evaluation) is that it gives agencies the potential to keep track of progress as part of a process of continual assessment. So, ‘rather than addressing impact as a question to be answered only once, usually after the intervention has been made, the relevance and usefulness of services should be continually assessed. Through this learning process, the organisation can adapt its services to better meet the needs of users’ (Johnson & Rogaly, 1999). Currently staff do not have the resources for covering impact, and incentives lean toward a focus on the input and output levels. This is particularly worrying in long-term or chronic emergencies where ample opportunities exist to develop relevant monitoring capacities, but where performance has generally proven to be little better than in short-term interventions.

A focus on impact in turn will have implications for the way in which humanitarian actors set the objectives of their interventions. Most intervention objectives are currently couched in terms of outputs or at best outcomes – and, as noted, are assessed as successful because they have achieved satisfactory results at this level. This practice is questionable and appears to be supporting conceptual problems with

LRRD, which in turn contributes to the poor performance on supporting sustainability/connectedness highlighted in Chapter 3. A shift to include impact monitoring will necessitate a re-examination of potential humanitarian action results, and support a realistic focus on what is achievable in terms of LRRD.

#### Shifting to a Balanced Quantitative and Qualitative Focus

Monitoring data should include both its current quantitative approach but complement this with a qualitative assessment. One of the main strengths of a reporting system that is 'real-time' is that it can establish why particular results are achieved – who the main actors are, who benefits, and who loses. If monitoring data can explain why an intervention has achieved particular results it may be more likely that this data will be useful and used.

ALNAP's research has not provided a sufficient empirical basis to judge the exact scale and nature of the trade-offs that each staff member must deal with, and thereby suggest what an optimal use of resources would look like. This topic will be explored further in the coming year as ALNAP looks more deeply into fieldlevel learning. The challenge is to be able to build the most appropriate mix of methodological techniques into the programme planning process, and to ensure that incentives are in place to encourage staff to use both quantitative and qualitative approaches. This would, however, require a substantial commitment by agencies in terms of training.

#### Consultation with Primary Stakeholders and Downward Accountability

There is a need for agencies to find an improved balance between reporting for upward accountability and downward accountability to primary stakeholders. There are good grounds for presuming that various interagency accountability initiatives aimed at improving downward accountability have not made a real difference. For example, over the three years of the Annual Review, 86 per cent of evaluations were rated as unsatisfactory in terms of consulting with primary stakeholders.

There is conceptual and policy agreement that consultation is both ethical and necessary. What is needed now is guidance on what is a minimum adequate level of consultation – what is feasible in different situations, how much it will cost in time and resources, and the necessary methods. The ALNAP Global Study on participation is well placed to provide guidance in this area.



## An Agenda for Improvement

### 4.5

Chapter 2 concludes with a set of recommendations for operational agencies, donors, and networks in response to the points raised above, with a focus on simplifying and harmonising systems and promoting downward accountability. Perhaps the most important recommendation for the sector as a whole is that a ‘community of practice’ be formed to continue dialogue on this issue and seek ways to implement the recommendations made in Chapter 2. It should be remembered that Chapter 2 constitutes a first, preliminary assessment of monitoring across the humanitarian sector. Having identified the key problems and constraints, highlighted good practice and made some suggestions for change, it is now up to key actors to take this agenda forward.



Section M

Meta-Evaluation

## Overview

M1

The purpose of a meta-evaluation is to assess the quality of a thematic group of evaluations, and subsequently to suggest improvements. As Lipsey points out (2000:212):

Meta-analysis and other forms of systematic synthesis of evaluation studies provide the information resources for a continuous improvement program for evaluation practice itself. By examining the patterns and relationships revealed by meta-analysis, an evaluator will better understand what program characteristics, outcome domains, and research methods are most likely to be important for a particular evaluation effort. As new evaluation studies are completed and added to cumulative syntheses, the knowledge resources of the evaluation field will become richer and more differentiated and their potential contribution to practice, in turn, will become more useful.

As in the two previous Annual Reviews, analysis is based on an assessment of the preceding year's set of evaluations against the ALNAP Quality Proforma (QP), which this year involved 34 English language reports and five French language reports.

Evaluation of Humanitarian Action (EHA) has many strengths. In particular, it allows us to look closely at key management issues, assess performance against some of the DAC criteria (including effectiveness, sustainability/connectedness and relevance/appropriateness), and understand better the contextual background of a particular emergency. This year good practice was found in particular in the reports by WFP and the Disasters Emergency Committee (DEC). While it is difficult to draw definite conclusions there has certainly been an improvement in the focus on longer term results. This is illustrated by a solid minority of reports achieving good rating against the DAC criteria 'impact'. However, this year's analysis also points to five generic weaknesses:

- 1 failure to use agency policies for evaluation purposes;
- 2 lack of attention to rights-based approaches (including gender equality) and

protection (evaluators clearly feel more comfortable dealing with management issues or the effectiveness of sectoral interventions than with rights-based issues);

- 3 questionable credibility of many reports due to inadequate methodology and/or because it is unclear from where conclusions are drawn;
- 4 failure to consult with primary stakeholders and/or to adequately describe the nature of this consultation;
- 5 recommendations that are poorly developed and therefore unlikely to be followed.

The last three points may be reasons, among others, why evaluation results are not being picked up on more fully.

This section begins with an introduction to the revised ALNAP QP (reproduced at the end of this chapter) and the assessment process. It then presents findings as organised by the QP headings. There is a general conclusion that, along with Box M1, elucidates a suggested new approach to EHA as recommended in the report on DEC agency interventions after the Gujarat earthquake (DEC, 2001). A year-on-year cumulative comparison for the three years covered by the ALNAP Annual Reviews is included as Box M3.

## The ALNAP Quality Proforma (QP)

### M1.1

The ALNAP QP was developed in 2000 by drawing on what was commonly accepted as good practice in EHA and evaluation in general. It is a 'live' document and continues to evolve as EHA evolves. The QP is not used to rank evaluation reports and no composite rating of individual reports is provided. Rather, the intention is to reach general conclusions on trends as well as strengths and weaknesses in EHA. Assessments using the QP are made entirely on the basis of information contained in an evaluation report; issues related to recommendations and follow-up are not covered unless discussed specifically in the report. As in the previous two years a 'satisfactory' rating is taken as the benchmark for adequate performance (as set out in the Guidance Notes column in the QP).

In the light of previous use in the Annual Review, the QP was revised this year to strengthen analysis. The main revisions were as follows:

- reorganisation for ease of use;
- addition of a section dealing with planning and implementation;
- addition of new Areas of Enquiry relating to: the cost of the evaluation (1.i); evaluator bias (2.v); and protecting confidentiality and promoting respect for stakeholders' dignity and self-worth (2.vi) (the latter areas are included in the American Evaluation Association The Program Evaluation Standards, 1994, and all three areas should be seen as central elements in EHA);
- separate ratings for each of the DAC/OECD criteria to facilitate a disaggregated understanding of evaluation performance against each criterion;
- clarification of some guidance notes, including addition of information in several areas to support determination of what can be considered 'satisfactory'; and removal of the C+ rating ('close to satisfactory') which was considered no longer necessary.

These clarifications and revisions have 'raised the bar'; the requirements to achieve a satisfactory rating have been made more stringent as QP guidance notes have been clarified. Some of the decline in reports' performance this year may be accounted for by these changes. This is noted where relevant in the text.<sup>1</sup>

## Assessment Process

## M1.2

In order to increase rigour and counter the potential for assessor bias and error, the assessments were undertaken by two assessors: the author of this chapter and Peter Wiles. Both were involved in the last two meta-evaluation exercises and the subsequent QP revision. Reports in French were rated by one assessor, Sylvie Robert.<sup>2</sup>

The assessment process for the reports in English was twofold. An initial assessment of the core evaluation reports was undertaken independently by each assessor.

Discussion on issues of interpretation of guidance notes, possible errors and omissions ensued, and was followed by a final independent review by each assessor. The resulting 90 per cent consistency rate was deemed an acceptable margin for the purposes of this meta-evaluation. Where there was inconsistent rating, results are not included in the analysis.<sup>3</sup>

### Quality Proforma Follow-up

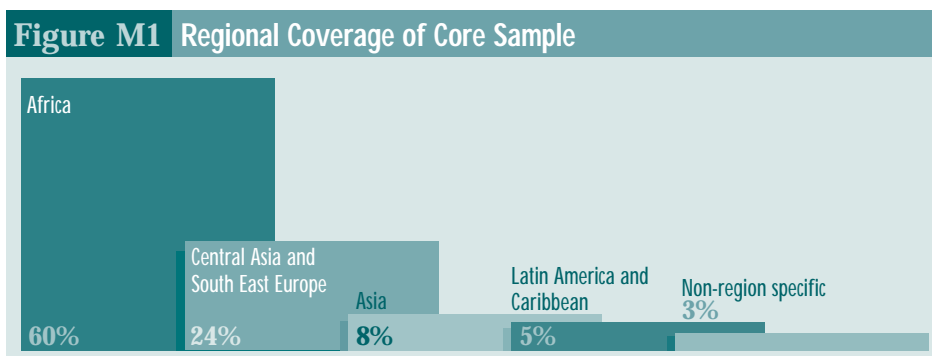
M1.3

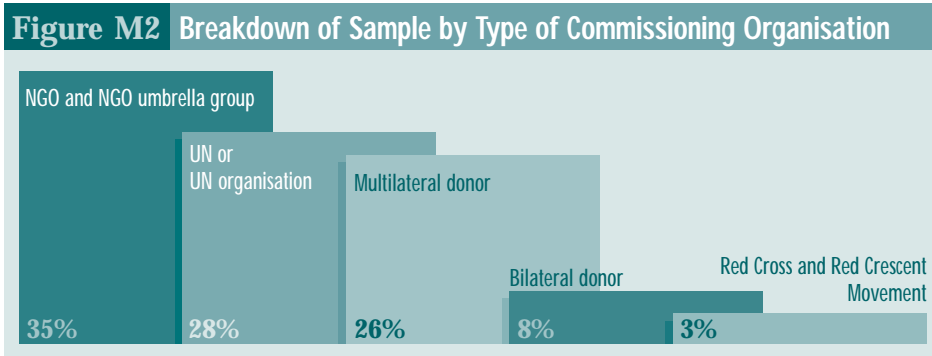
The quality assessments completed for Annual Review 2002 were returned to relevant agencies, but elicited almost no response. Thus a more proactive approach has been adopted this year whereby one of the assessors will be available to discuss individual assessments with commissioning agencies.

### The Sample

M1.4

Thirty-seven reports cover 24 countries or regions. From Africa: Angola (2 reports); Burundi (4 reports); DR Congo (4 reports); Ethiopia (2 reports); Great Lakes; Kenya; Liberia; Mozambique; Sierra Leone (4 reports); Somalia; Sudan; and Uganda. From Central Europe/Asia: Azerbaijan; Afghanistan (3 reports); Bosnia; northern Caucasus; Croatia; Kosovo; and Iran. From Asia: Bangladesh; India and DPR Korea. From Latin America: Brazil and Colombia. The remaining two reports had a multi-country focus (see Figure M1).





Fourteen of the reports were commissioned by NGOs (of which five were reports on Oxfam); 11 by the UN system (of which eight were on WFP); 10 by ECHO; three by bilaterals; and one by the IFRC (see Figure M2).

## Timeliness

## M1.5

In order to facilitate the use of results and recommendations, an analysis was carried out of whether evaluations were completed and reports published in a timely fashion. Thirty-six reports provided data sufficient to facilitate this analysis, although this data was in many cases incomplete or ambiguous. For example, many reports did not specify when the intervention began and ended or the phase of the intervention being evaluated.

The following illustrates that EHA is being carried out in a timely manner: 16 evaluations were conducted on ongoing interventions; 12 evaluations were conducted within one month of completion of operations; four evaluations were conducted within two to three months after completion of operations; and two evaluations were conducted more than three months after completion of operations. The large number of evaluations carried out during ongoing operations is probably due to the long-term nature of many of the interventions included in the 2002 set.

For the most part reports were produced in a timely fashion, with 13 reports finalised within one month of the evaluation, 11 within two to three months, and eight within four to seven months. While we do not have comparative figures for



other sectors, this performance seems at least satisfactory as far as promoting the use of reports is concerned.

### Three-year Comparative Analysis

### M1.6

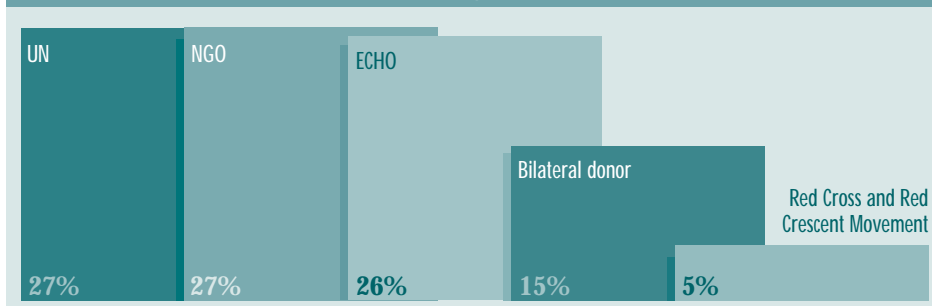
A new feature of this year's meta-evaluation is a year-on-year comparative analysis based on comparable Areas of Enquiry from the QP used in 2001, 2002 and 2003. This comparative analysis involves a total of 127 reports.

As a slightly different rating system has been used over the three years, analysis has been carried out using 'satisfactory' and 'unsatisfactory' ratings only – as defined in the Guidance Notes of the QP.

The breakdown of the sample by agency over three years is given in Figure M3. Between them the UN, NGOs and ECHO constitute 77 per cent of the evaluations included in the meta-analysis in the Annual Reviews over this and the last two years.

The limited number of evaluations commissioned by bilaterals provided to ALNAP each year suggests that Collinson & Buchanan-Smith's (2002) analysis concerning lack of accountability of donors vis-a-vis their willingness to commission independent evaluations holds true. ICRC, with two evaluations provided over three years, is also poorly represented.<sup>4</sup>

**Figure M3** Breakdown of Agencies Included in QP Assessment for 2001, 2002 and 2003



While the analysis of EHA shows some strengths, for example, in application of some of the DAC criteria, attention to rights-based issues and consultation with and participation of primary stakeholders stand out as particular weaknesses.

## Assessment Against the ALNAP Quality Proforma: 2002 Reports

M2

Information included in the Area of Enquiry column in the tables below provides the outline of the area being considered. Further details as to how the rating was determined can be found in the Guidance Notes column in the QP.

### Proforma Section 1: Evaluation Terms of Reference (TOR)

M2.1

Note: Of the 39 reports assessed, 11 included no TOR in the version received. The analysis below and figures in the tables are therefore based on the 28 reports that included TOR, except for QP area 1.2i.

#### TOR Focus and Use (1.1ii; 1.1iii; 1.1iv; & 1.1vi)

In terms of the statement on the intervention to be evaluated, reports were required to include adequate details on the emergency context, intervention objectives and key stakeholders involved. Only nine reports that included a TOR managed to provide adequate information in all three areas. Of the remaining 19, details concerning key stakeholders were least often provided (in only four of the 19 reports); better information was included on context and objectives (provided in nine of the 19). Failure to set out clearly in the TOR how primary stakeholders should be consulted may be one of the reasons for poor consultation with and participation of primary stakeholders in the evaluation process (see M2.6 later).

Commissioning agencies should, as a matter of course, include in the TOR a requirement that evaluators consult adequately with primary stakeholders. What can be considered 'adequate' is:

**Table M1** TOR Focus and Use

| Area of Enquiry   | Rating         | % of Reports Attaining Rating (rounded) |
|---|----------------|---|
| <b>1.1ii</b>  |                |   |
| Quality of TOR statement on the intervention to be evaluated.                     | Good           | 4                                       |
|   | Satisfactory   | 32                                      |
|   | Unsatisfactory | 56                                      |
|   | Poor           | 8                                       |
| <b>1.1iii</b>   |                |   |
| Quality of TOR statement on the purpose of the evaluation.                        | Good           | 26                                      |
|   | Satisfactory   | 52                                      |
|   | Unsatisfactory | 8                                       |
|   | Poor           | 14                                      |
| <b>1.1iv</b>  |                |   |
| Quality of TOR statement on the primary focus of the evaluation.                  | Good           | 26                                      |
|   | Satisfactory   | 70                                      |
|   | Unsatisfactory | 4                                       |
|   | Poor           | 0                                       |
| <b>1.1vi</b>  |                |   |
| Quality of TOR statement on intended use and user(s) of the evaluation output(s). | Good           | 4                                       |
|   | Satisfactory   | 11                                      |
|   | Unsatisfactory | 63                                      |
|   | Poor           | 22                                      |

- that sufficient information is gained from primary stakeholders, including from both sexes and different ethnic groups, to allow conclusions to be formulated about the intervention;
- that primary stakeholders be given an opportunity to be active participants in the evaluation process, even if only through focus groups or PRA exercises;
- that primary stakeholders' perspectives' can be triangulated with those of other key stakeholders.

Reports were strong on providing background information on the purpose and primary focus of the evaluation. Of the 26 reports that covered this area, 17 noted a joint lesson learning and accountability focus, six a lesson learning and three an accountability focus – though for the most part reports did not specify the relative emphasis placed on each. The constraints to achieving both lesson learning and accountability functions from the same evaluation are discussed in Annual Report 2002. It is therefore possible to conclude that commissioning agencies appear to be including both purposes in a rote manner without considering the consequences of this for evaluation process and use. Of the 23 reports noting primary focus, nine had a programme focus (mainly ECHO reports), five had a project focus<sup>5</sup>, and the remainder had either a policy, institutional or joint focus.

One of the weaker areas of the reports (1.1vi) is the extent to which they outline the intended uses of the evaluation findings. Sixty-three per cent of reports rated as unsatisfactory and 22 per cent as poor in this area, with no mention of this topic. Furthermore, commissioning agencies are not following up on their responsibility to ensure that evaluation results are used, despite widespread acknowledgement in many cases that reports do not receive sufficient attention. The one good practice example this year is the WFP evaluation of its intervention in the Great Lakes region (September 2002, Annex 1) which notes that: '[T]he report will be presented to WFP's Executive Board; key recommendations arising from the evaluation will be used in the preparation of a Management Response Matrix which will outline how the WFP Regional Bureau in Kampala intends to follow up on the evaluation's key findings and recommendations; and [there will be] dissemination through WFP's website and a publicly available summary.'

#### TOR process and team make-up: evaluation cost (1.1i)

This is an area where there has been surprisingly little analysis in EHA, and where no interagency standards exist. There does not even appear to be a requirement for commissioning agencies to report on likely costs of evaluations in the TOR. Thus a new Area of Enquiry was added this year, and it was found that only two reports included the cost of the evaluation (Norway Ministry of Foreign Affairs, November 2001; WHO, December 2002). In the former case the cost was some US\$175,000; in the latter it was US\$20–30,000 out of a total expenditure of US\$1.7m in 2002 – or between some one-and-a-half and 2 per cent of total WHO expenditure. This area should be included in TOR for transparency and cost-effectiveness reasons, and to allow an assessment of whether agencies are

**Table M2** TOR Process and Team Make-Up

| Area of Enquiry   | Rating   | % of Reports<br>Attaining<br>Rating (rounded) |
|---|--|---|
| <b>1.1i</b><br>Cost of the evaluation.  | Good<br>Satisfactory<br>Unsatisfactory<br>Poor | 4<br>4<br>0<br>92                             |
| <b>1.1v</b><br>Quality of TOR statement on the expectation of good practice in approach and method.   | Good<br>Satisfactory<br>Unsatisfactory<br>Poor | 4<br>0<br>86<br>11                            |
| <b>1.1vii</b><br>Quality of TOR guidance on the evaluation report format.   | Good<br>Satisfactory<br>Unsatisfactory<br>Poor | 29<br>11<br>7<br>54                           |
| <b>1.1viii</b><br>Evaluation Timeframe <i>a. Timeliness</i><br>The TOR should outline the rationale for the timing of the evaluation.   | Good<br>Satisfactory<br>Unsatisfactory<br>Poor | 15<br>25<br>0<br>60                           |
| <b>1.1viii</b><br>Evaluation Timeframe <i>b. Sufficiency</i><br>Sufficient time should be allowed to develop methods; review background/contextual information; carry out fieldwork; undertake analysis at all stages of the evaluation; and finalise the report. | Good<br>Satisfactory<br>Unsatisfactory<br>Poor | 0<br>50<br>21<br>29                           |
| <b>1.1ix</b><br>Quality of TOR clarification process.   | Good<br>Satisfactory<br>Unsatisfactory<br>Poor | 4<br>11<br>4<br>81                            |
| <b>1.2i</b><br>Nature, make-up and appropriateness of the evaluation team.  | Good<br>Satisfactory<br>Unsatisfactory<br>Poor | 8<br>8<br>38<br>46                            |

spending an adequate percentage of an intervention budget on evaluation where agency standards do exist.<sup>6</sup>

#### Expectation of good practice in approach and method (1.1v)

In this area, TOR are expected to outline application of DAC criteria; reference international standards, including international law; talk about the importance of a multi-method approach; explain the consultation process with key stakeholders, including primary stakeholders; and bring in the key issue of gender analysis. Only one report, the DEC evaluation of interventions after the Gujarat earthquake (DEC, December 2001), managed to cover all these areas adequately.

It is useful to consider the disaggregated breakdown of this section. Almost no TOR required evaluators to either use international standards, such as the Sphere standards, or to examine whether the intervention had used international standards during its implementation. Protection was similarly largely ignored in the TOR as an issue to be evaluated. On the other hand, TOR generally set out a requirement to use the DAC criteria and explained what the criteria meant (in 19 out of 28 reports); a majority of TOR also required attention to gender equality (16 out of 28 reports). The requirement to develop a multi-method approach and consult with key stakeholders was weaker (11 out of 28 reports).

In the light of this disaggregated breakdown, there is a fairly clear correlation between requirements in the TOR and what was actually done by evaluators, particularly in terms of the lack of attention to international standards and protection and the relative success in application of the DAC criteria. Commissioning agencies and evaluators may want to mull this over when formulating and finalising their TOR. The area of clarification of TOR between the commissioning agency and evaluation team (1.1ix) is one that we hear almost nothing about in evaluation reports, with 81 per cent of reports with TOR making no mention of this topic despite its importance to the overall direction of the evaluation.

#### Evaluation timeframe (1.viii.a & b)

The evaluation timeframe is also inadequately reported in terms of the reason why the evaluation is being carried out at a particular time (60 per cent of evaluations rated as poor); however, reporting on whether sufficient time had been allowed for the evaluation was better (50 per cent rated as satisfactory).

### The nature, make-up and appropriateness of the evaluation team (1.2i)

This area is important for establishing the credibility of the evaluation process, yet only four reports managed to do this (DEC, December 2001; Norway Ministry of Foreign Affairs, November 2001; Oxfam, September 2001; and WHO, December 2002). Forty-six per cent of the 28 reports with TOR provided no information on this area at all.

Of particular interest here is the make-up of evaluation teams which is included in, or could be extrapolated from, reports. Of the 61 evaluators employed, 46 were expatriates, seven were locally based consultants (usually citizens of the countries where the evaluation took place), and eight were agency staff members (from two Oxfam evaluations). While this is a better balance than that noted in Annual Review 2001 for the Kosovo evaluations, where 52 of 55 consultants were expatriates, this still represents a serious imbalance and under-employment of locally based professionals. Only the DEC has evidenced consistent good practice in this area, this year hiring a team with a complementary mix of international and locally based consultants (DEC, December 2001). In addition, of the 61 evaluators, 10 were internal agency staff. However, the implications of this for evaluation practice are not discussed.

One of the key cross-cutting themes in Chapter 3 this year was the general lack of capacity building in humanitarian action. The same could be said for EHA; agencies could usefully maintain and use rosters of locally based evaluators and attempt to ensure a better mix of international and locally based evaluators. ALNAP could also play a central role here, both in terms of developing national evaluation capacity through training and maintaining a roster of consultants.<sup>7</sup>

## Proforma Section 2: Evaluation Approach and Methods

### M2.2

#### Overview

Description of the evaluation approach did not display any systematic good practice, and report methodology sections in general did not provide a basis from which it was possible to assess the likely accuracy of findings. This may be because evaluators do not feel it is necessary to elaborate on the methodology used. However, this lack is specific to EHA; in both mainstream evaluation practice and the evaluation of development interventions, greater attention is given to establishing the credibility

**Table M3** Appropriateness of Evaluation Approach

| Area of Enquiry   | Rating         | % of Reports<br>Attaining<br>Rating (rounded) |
|---|----------------|---|
| <b>2.i</b>  |                |   |
| Appropriateness of the overall evaluation approach.                           | Good           | 0   |
|   | Satisfactory   | 5   |
|   | Unsatisfactory | 8   |
|   | Poor           | 87  |
| <b>2.ii</b>   |                |   |
| Appropriateness of the evaluation methods selected.                           | Good           | 3   |
|   | Satisfactory   | 0   |
|   | Unsatisfactory | 69  |
|   | Poor           | 28  |
| <b>2.iii</b>  |                |   |
| Appropriateness of the planned application of the DAC criteria and rationale. | Good           | 5   |
|   | Satisfactory   | 8   |
|   | Unsatisfactory | 13  |
|   | Poor           | 74  |
| <b>2.iv</b>   |                |   |
| Consideration given to constraints.   | Good           | 0   |
|   | Satisfactory   | 19  |
|   | Unsatisfactory | 42  |
|   | Poor           | 39  |
| <b>2.v</b>  |                |   |
| Consideration given to evaluator bias.  | Good           | 0   |
|   | Satisfactory   | 3   |
|   | Unsatisfactory | 5   |
|   | Poor           | 92  |
| <b>2.vi</b>   |                |   |
| Consideration given to confidentiality and dignity.                           | Good           | 3   |
|   | Satisfactory   | 3   |
|   | Unsatisfactory | 7   |
|   | Poor           | 87  |
| <b>5.3i</b>   |                |   |
| Quality of application of the selected evaluation methods.                    | Good           | 3   |
|   | Satisfactory   | 5   |
|   | Unsatisfactory | 76  |
|   | Poor           | 16  |



of evaluation methodology. Experienced evaluators know that when evaluations bring unwelcome findings and recommendations the first thing that may be questioned is the evaluation methodology.

Producing a credible description of methodology would, in the majority of cases, require only a little more effort – for example, noting the numbers of primary stakeholders consulted, broken down by sex and other salient social characteristics; where and when they were consulted; and the methods used for consultation (survey questionnaire, focus group, etc). That the majority of reports do not report even such basic information suggests sloppy practice on the part of evaluators and commissioning agencies, and undermines their credibility.

#### Appropriateness of the overall evaluation approach (2.i)

Reports were assessed concerning the extent to which the overall evaluation approach was clearly outlined and the appropriateness of choice established relative to the evaluation's primary purpose, focus and end-users. 'Approach' here means the wider conceptual framework used and the evaluation tradition being drawn upon, such as accountability oriented, utilisation-focused, or empowerment evaluation approaches.

Only two reports were assessed as satisfactory in this area: the evaluation of the DEC's intervention after the Gujarat earthquake (DEC, December 2001) and the evaluation of the Norwegian Red Cross (Norwegian Ministry of Foreign Affairs, November 2001). Eighty-seven per cent of the set rated as poor. In the DEC case the rationale for primary stakeholder consultation is clearly set out and the community survey technique used is close to the empowerment evaluation approach. The Norwegian Red Cross evaluation analyses the way in which definition of the term 'humanitarian' affected the evaluation methodology. In both cases there was an explicit discussion of why a particular evaluation approach was taken.

The implications of the lack of attention to wider evaluation discourse were discussed in detail in Annual Review 2002, and include: a lack of conceptual direction for the evaluation; an inability to rationalise why a particular evaluation methodology has been selected; a fall-back on 'standard' evaluation techniques with little experimentation; and lack of attention to causality.

### Appropriateness of the evaluation methods selected and appropriateness of planned application of the DAC criteria (2.ii & iii)

A large majority of reports were rated as unsatisfactory or poor in both these areas. In terms of rating of appropriateness of selected evaluation methods, reports most often missed reference to international standards, including international law, and gender analysis. Among those reports rated satisfactory or better was the evaluation of WFP's intervention in Iran supporting Iraqi and Afghan refugees (WFP, September 2002b). This report discusses in some detail the reason for selection of particular camps to visit and the random selection of refugees for interview purposes. It provides a checklist for refugee camp visits and notes the methods used – in particular, focus groups. It is also gender sensitive and notes the use of control groups.

Use of control groups is a rare phenomenon in EHA, and it is common for reports to argue that they cannot conclude whether results were caused by the intervention because of difficulties of attribution. For example, World Vision (June 2001:5) notes: '[W]hile the emergency interventions undoubtedly played an important role in reducing the levels of malnutrition, a number of other factors may have contributed to this reduction.' Use of a rudimentary control group approach, for example, interviews with a small sample of the affected population who have not been included in the intervention target group, can help overcome this problem. Not surprisingly, because a majority of reports were assessed as unsatisfactory in selection and detailing of evaluation methodology a similar number were assessed as unsatisfactory in the use of evaluation methods (5.3i).

The UNHCR (May 2002) evaluation of the protection of children, not included in this year's meta-analysis because of its thematic focus, is in many respects an example of good practice in the application of methodology. In particular, the report has included an Annex on lessons learned in the evaluation process and methodology. These lessons include: the need for triangulation between different sources of data; the need for the UNHCR evaluation office to communicate the important balance between 'independence' and 'internal purview' when announcing the evaluation; the importance of having a representative and active steering committee to guide the evaluation; the importance of having briefing and debriefing sessions with the country office; and the key role that focus groups can play in providing qualitative information.

In terms of the application of DAC criteria, reports were required to address each of seven criteria to achieve a satisfactory rating. Seventy-four per cent of reports did not discuss the criteria at all in the methods section, except in a very general fashion and even where attention to the criteria was included as a requirement in the TOR. Of those that did include a discussion, 'coherence' was the criterion most often missed.<sup>8</sup>

#### Consideration given to evaluator bias; and stakeholder confidentiality and dignity (2.v & 2.vi)

Bias is a well-known feature of evaluations, both conceptual bias that inevitably comes with any individual's perspective and/or bias that is introduced as a result of an evaluator having been associated with an intervention. One of the key points of a well-designed methodology is to guard against bias. Evaluations did not see fit to address this topic. For example, seven evaluations involved staff from the agency being evaluated as part of the evaluation team. However, there is no discussion in the reports as to why a staff member was included and, for example, the potential bias or benefits this brings.

Ensuring confidentiality and the dignity of key stakeholders should be central to any evaluation, but is particularly important in EHA where primary stakeholders may be at risk and where they have often been subject to trauma. Doubtless evaluators treat primary stakeholders with respect and would never dream of putting them at risk intentionally; it is important to acknowledge this in their reports. In a similar fashion, evaluation reports should point out how the views of other key stakeholders are kept confidential and how the evaluation method encouraged key stakeholders to express their independent opinions. Ethical research standards for interviews do exist, for example, providing interviewees with a form signed by the evaluators noting that their views will be kept confidential. Such a system could be usefully adapted for EHA.

An evaluation may be the first time primary stakeholders are listened to seriously (DEC, December 2001, Vol 3, Methodology: pt 6): 'Some members of the community stated that no one else had asked what they wanted or needed, or how they felt about the response.' This raises the question as to whether part of the purpose of an evaluation should be to give voice to the usually voiceless – as professed in empowerment evaluation and as the DEC report argues. Not all evaluators would agree with this perspective but rather consider that evaluation should be an

'objective' exercise. This is why providing a rationale for the evaluation approach and details on potential evaluator bias is crucial.

## Assessing Contextual Analysis

M2.3

### Overview

The context section of an evaluation report should constitute a concise and relevant background that allows the reader to understand the situation in which the intervention took place, and how context is relevant to, and has been taken into consideration in, the evaluation. Understanding of context is important for evaluation purposes primarily because it supports attribution of results. This year's reports rated fairly well in attention to context. To receive a satisfactory rating reports were required to include relevant details on historical, social (including gender analysis), economic, political, and cultural features. Reports tended to focus on providing background to the sector being assessed, and most often left out historical and cultural features.

### Analysis of context (3.i)

In terms of analysis of the affected area and population there was significant good practice, for example USAID (April 2001), WFP (January 2002; April 2002) and ECHO (October 2001b). The WFP Angola evaluation (April 2002) section on context includes details on the war and agriculture and the impact of this on food aid and the food economy; coping strategies of primary stakeholders (including trade in semi-urban settlements); the economy; geographical location of insecurity (including a map); gender and poverty; land tenure; and the lack of government policy on LRRD. All of these areas are of relevance to the intervention and are drawn upon in the analysis of results.

### Quality of use of context information (5.1i)

Fewer reports integrated the discussion of context into the analysis, and 63 per cent of reports were rated as unsatisfactory or worse in this area - although the tendency to provide stand-alone contextual sections with little reference to the rest of the report had diminished. On the other hand about 80 per cent of those reports that did include a satisfactory or better context section also managed to integrate this with discussion of results - i.e., to demonstrate how context affected achievement.

**Table M4** Quality of Contextual Analysis

| Area of Enquiry  | Rating         | % of Reports Attaining Rating (rounded) |
|--|----------------|---|
| <b>3.i</b>   |                |   |
| Quality of the evaluation's analysis of context.   | Good           | 16                                      |
|  | Satisfactory   | 32                                      |
|  | Unsatisfactory | 39                                      |
|  | Poor           | 13                                      |
| <b>3.ii</b>  |                |   |
| Quality of the evaluation's analysis of past involvement of the agency and its local partners. | Good           | 7                                       |
|  | Satisfactory   | 29                                      |
|  | Unsatisfactory | 32                                      |
|  | Poor           | 32                                      |
| <b>3.iii</b>   |                |   |
| Quality of the evaluation's analysis of the crisis to which the intervention is responding.    | Good           | 9                                       |
|  | Satisfactory   | 15                                      |
|  | Unsatisfactory | 43                                      |
|  | Poor           | 33                                      |
| <b>5.1i</b>  |                |   |
| Quality of the use made by the evaluation of contextual information.                           | Good           | 14                                      |
|  | Satisfactory   | 23                                      |
|  | Unsatisfactory | 49                                      |
|  | Poor           | 14                                      |

### Agency involvement (3.ii)

Details of the past involvement of the agency and its partners in the geographical area of the intervention is less well covered. Indeed it might appear that the intervention occurred in a historical vacuum in many of the reports, whereas the prior involvement of an agency in an area has been identified as a key factor in success (see Annual Review 2002). That one-quarter of evaluation reports did not see fit to cover this area, and a further 43 per cent were rated as unsatisfactory, suggests that the importance of prior involvement and building partnerships which can be drawn on in an emergency is not sufficiently recognised in EHA.

## Assessing the Analysis of the Intervention

## M2.4

### Evaluation of policies and principles (4.1i)

Reports were generally weak in terms of evaluating adherence to policies. Reports commissioned by ECHO and NGOs tended to be weaker in this area, with only one report for each type of agency rating as satisfactory (ECHO, December 2001f; DEC, December 2001). Reports commissioned by UN agencies were stronger, with four (three WFP and one UNMAS, February 2002) reports rated as good, and two (both WFP) as satisfactory. Each of these WFP reports tells the story of how the introduction of the WFP policy From Crisis to Recovery influenced the planning and implementation of the respective country programmes; the Great Lakes report (September 2002) goes further and analyses the relation between government policy and WFP's intervention.

It is surprising that more evaluations do not use agency policy as a standard against which results can be measured given the key role that such policy should play in guiding agency action. This is partly explained by TOR not including this as a requirement.

**Table M5** Institutional Considerations

| Area of Enquiry   | Rating         | % of Reports Attaining Rating (rounded) |
|---|----------------|---|
| <b>4.1i</b>   |                |   |
| Quality of the evaluation of agency guiding policies and principles.              | Good           | 12                                      |
|   | Satisfactory   | 17                                      |
|   | Unsatisfactory | 21                                      |
|   | Poor           | 50                                      |
| <b>4.1ii</b>  |                |   |
| Quality of the evaluation of an agency's management and human resource practices. | Good           | 18                                      |
|   | Satisfactory   | 37                                      |
|   | Unsatisfactory | 30                                      |
|   | Poor           | 15                                      |

### Management and human resource practices (4.1ii)

As in the two previous Annual Reviews, this area has proven to be a strength with 55 per cent of reports this year rating as satisfactory or good. Reports contained most information about staff turnover and field/HQ relations; less information was provided on briefing and debriefing procedures and training.

There were a number of examples of good practice where reports made the connection between management practices and intervention results. Among these was the assessment of CARE's programme in Afghanistan (CARE, September 2002) which includes extensive analysis of field/HQ communication; the implications of a lack of training on intervention impact; lack of learning from earlier emergencies (e.g., around the importance of income generation schemes in the recovery phase); and how security issues affected programming. Another good practice case was the evaluation of the IFRC intervention after the Goma volcanic eruption (IFRC, September 2002) This provided a good assessment of intra-agency communication/coordination, including dispatch of emergency teams; intra-Federation communication; level of preparedness of Federation staff; the level of experience of emergency teams; and training.

### Overview

This section of the QP was one of those areas revised this year to reflect more accurately the various stages of the project cycle, from planning to monitoring. Overall, evaluation of the project cycle process was a relatively strong area of assessment with reports rated as satisfactory or better in about 50 per cent in three areas: evaluation of implementation, monitoring, and expenditure. Reports tended to be consistent in their coverage of project planning and implementation – that is, they were either rated as satisfactory or better, or unsatisfactory or worse, in all areas of enquiry. WFP reports were particularly strong in this area.

### Evaluation of needs and livelihoods assessment (4.2i)

Forty per cent of reports were attuned to this issue and included an analysis of both whether the intervention had carried out an adequate needs assessment, and the importance of this. Several reports (e.g., World Vision, June 2001; DEC, December 2001; WFP, September 2002b) go beyond a critique of the lack of an adequate needs assessment and include an analysis of the ways in which livelihood strategies could have been more adequately covered in the intervention planning process.

**Table M6** Needs Assessment, Objectives, Planning and Implementation

| Area of Enquiry  | Rating         | % of Reports<br>Attaining<br>Rating (rounded) |
|--|----------------|---|
| <b>4.2i</b>  |                |   |
| Quality of evaluation of the needs and livelihoods assessments(s) that informed the intervention.  | Good           | 8   |
|  | Satisfactory   | 32  |
|  | Unsatisfactory | 30  |
|  | Poor           | 30  |
| <b>4.2ii</b>   |                |   |
| Quality of evaluation of the intervention objective(s).  | Good           | 22  |
|  | Satisfactory   | 19  |
|  | Unsatisfactory | 47  |
|  | Poor           | 12  |
| <b>4.2iii</b>  |                |   |
| Quality of evaluation of the intervention planning processes (including design).   | Good           | 11  |
|  | Satisfactory   | 22  |
|  | Unsatisfactory | 62  |
|  | Poor           | 5   |
| <b>4.2iv</b>   |                |   |
| Quality of evaluation of the intervention implementation process.  | Good           | 10  |
|  | Satisfactory   | 39  |
|  | Unsatisfactory | 51  |
|  | Poor           | 0   |
| <b>4.2va</b>   |                |   |
| Quality of evaluation of monitoring and/or real-time evaluation mechanisms.  | Good           | 19  |
| • Analysis of the intervention's monitoring and/or real-time evaluation mechanisms and the effect on intervention results.   | Satisfactory   | 44  |
|  | Unsatisfactory | 25  |
|  | Poor           | 12  |
| <b>4.2vb</b>   |                |   |
| Quality of evaluation of monitoring and/or real-time evaluation mechanisms. • Assessment of the indicators used. Where the intervention activities span relief, rehabilitation and/or development, indicators should be evaluated relative to each type of activity. | Good           | 0   |
|  | Satisfactory   | 41  |
|  | Unsatisfactory | 24  |
|  | Poor           | 35  |
| <b>4.2vi</b>   |                |   |
| Quality of evaluation of the intervention expenditure.   | Good           | 18  |
|  | Satisfactory   | 32  |
|  | Unsatisfactory | 32  |
|  | Poor           | 18  |



For those reports rated as unsatisfactory the majority either only mentioned the livelihoods assessment in passing or included no details on primary stakeholder consultation and participation. Almost one-third of reports made no mention of this topic.

#### Evaluation of intervention planning processes (4.2iii)

This has proven a difficult area for EHA to come to grips with, with some 67 per cent of evaluations (25 reports) rated as unsatisfactory or poor. Of these 25 reports, 19 did not include adequate details on primary stakeholder consultation and participation in the planning processes; as was seen in Chapter 3, facilitating consultation and participation in planning is one of the most difficult areas of humanitarian action. By failing to pay systematic attention to this issue evaluators are compounding the problem.

#### Evaluation of the intervention implementation processes (4.2iv)

Evaluation of implementation processes was satisfactory in about half of all reports. However, lack of attention to primary stakeholder participation and consultation was a significant problem. UN agencies rated higher than NGOs, suggesting that the perceived greater capacity of NGOs to foster consultation and participation of primary stakeholders has not translated into capacity to facilitate evaluation of this aspect of humanitarian action.

#### Evaluation of monitoring and/or RTE mechanisms, and indicators (4.2va & b)

Sixty-three per cent of reports in the former area were assessed as satisfactory or better. Of the six reports rated as good in relation to evaluation of monitoring, four were commissioned by WFP (April 2002, September 2002, September 2002b, December 2001), one by the Norwegian Ministry of Foreign Affairs (November 2001), and one by ECHO (December 2001c). The significant number of reports that did not cover monitoring adequately is surprising given that this is usually an area close to evaluators' hearts. That four reports barely touched on this issue suggests a major oversight by commissioning offices in their vetting of reports. In terms of attention to indicators, 59 per cent of reports did not cover this area even though the development of indicators is essential to results-based planning and should be a central feature of evaluators' approaches.

#### Quality of evaluation of the intervention expenditure (4.2vi)

Given the arcane budget codes of some agencies it is often difficult to assess whether funds are used for relief or rehabilitation – phases of a response which in any case

often merge into each other. As noted in the previous two Annual Reviews, agencies need to do a better job of delineating the different stages of their interventions, including financial allocations, so that relief and rehabilitation can be evaluated against appropriate indicators.

That 50 per cent of reports were assessed as good or satisfactory in this area would appear to be an improvement on the past two years, although because of differences in phrasing in the Proforma it is not possible to make direct comparison. Of the six examples of good practice, four were evaluations commissioned by WFP and two by NGOs. It is not surprising to find a sound level of detail in the WFP reports as part of their mandate was to investigate the LRRD continuum. In total, about half of the ECHO and NGO reports performed credibly.

## Consideration Given to Cross-cutting Themes

M2.5

### Evaluation of the intervention's adherence to international standards (4.3i)

As with last year's assessment, evaluation of how far the intervention adhered to international standards was again inadequate, with 86 per cent of reports rated as unsatisfactory or poor (the same as in Annual Review 2002). International standards such as the Sphere standards are either not being used by agencies or, less likely, this issue is not being picked up by evaluators. A parallel finding is that evaluators are not themselves using international standards as a means of evaluating interventions (5.3iii) - one of the reasons for this being that most EHA is organised around the OECD/DAC criteria to the seeming exclusion of other international standards. This in turn may be because most commissioning agencies do not require that these standards be used (see the analysis of Area of Enquiry 1.v on good practice in evaluation method, in Section M2.1 above). There is one innovative good practice example in this area, showcased in Box M1, which is the DEC's evaluation of NGO interventions after the Gujarat earthquake.

### Consideration given to coordination activities (4.3ii)

Results were satisfactory or better in 52 per cent of cases - not as positive as for 2002 when the figure was 67 per cent. Five reports were rated as good this year, including the evaluation of WFP interventions in Azerbaijan (April 2002a). This covers WFP's coordination mechanisms and relations with the national government, other UN agencies, and implementing partners. The majority of reports that include

**Table M7** Evaluation of Cross-cutting Themes

| Area of Enquiry   | Rating         | % of Reports<br>Attaining<br>Rating (rounded) |
|---|----------------|---|
| <b>4.3i</b>   |                |   |
| Quality of evaluation of the intervention's adherence to international standards.   | Good           | 3   |
|   | Satisfactory   | 11  |
|   | Unsatisfactory | 35  |
|   | Poor           | 51  |
| <b>4.3ii</b>  |                |   |
| Quality of evaluation of the consideration given to coordination activities.        | Good           | 16  |
|   | Satisfactory   | 36  |
|   | Unsatisfactory | 29  |
|   | Poor           | 19  |
| <b>4.3iii</b>   |                |   |
| Quality of evaluation of the consideration given to protection.                     | Good           | 4   |
|   | Satisfactory   | 4   |
|   | Unsatisfactory | 24  |
|   | Poor           | 68  |
| <b>4.3iv</b>  |                |   |
| Quality of evaluation of the consideration given to gender equality.                | Good           | 14  |
|   | Satisfactory   | 5   |
|   | Unsatisfactory | 38  |
|   | Poor           | 43  |
| <b>4.3v</b>   |                |   |
| Quality of evaluation of the consideration given to vulnerable/marginalised groups. | Good           | 12  |
|   | Satisfactory   | 24  |
|   | Unsatisfactory | 38  |
|   | Poor           | 26  |
| <b>5.3iii</b>   |                |   |
| Reference made to international standards.  | Good           | 3   |
|   | Satisfactory   | 5   |
|   | Unsatisfactory | 34  |
|   | Poor           | 58  |

**Box M1** Good Practice in the Use of International Standards? Using the Red Cross/Red Crescent Code of Conduct and Sphere Standards

The DEC evaluation of eight British NGO interventions after the January 2001 Gujarat earthquake included innovative use of the Red Cross/Red Crescent Code of Conduct as a measure of performance (December 2001:6):

We use the Red Cross Code as the basis from which to explore values because it is the most widely accepted set of humanitarian values and all DEC members must sign up to it ... The Code was evolved in the West and has not been negotiated with local NGOs or the people in need. In the decade since the Code was devised little has been done to promote it and too often it is just a 'badge' acquired easily by declaration ... but it is in the public domain, and anyone donating to the DEC or receiving its aid could reasonably expect agencies to follow it.

The intervention is then evaluated against the 10 sections of the Code, which are rated on points out of 10, with a cumulative rating given. The evaluation also includes an assessment against the Sphere standards in respect to training and DEC members' awareness of the standards, and the water, sanitation and shelter standards.

Without doubt the Code of Conduct is one standard against which humanitarian action should be assessed. However, the DEC report does not make the case as to why it should be the main standard and therefore why it should replace those evaluation mechanisms that are widely understood and in use, such as the OECD/DAC criteria. Evaluators have come to understand over the last decade that 'paradigm wars' as to the most effective evaluation approach are often not useful, and that the most effective evaluation approaches are those that use complementary methods. In any case there is considerable overlap between the several sections of the Code of Conduct and the OECD/DAC criteria (e.g., in relation to coherence, coverage, appropriateness/relevance and sustainability/connectedness) and most of the OECD/DAC criteria are covered in the DEC report. Lastly, while assigning scores to each of the sections of the Code could be useful, it would need more detailed discussion of how scores are to be assigned and whether any weighting should be given before this could be carried out comparatively.

Nevertheless the DEC innovation is certainly welcome and stands out in a field where there is very little experimentation.

an analysis of coordination issues tend to focus on sharing of knowledge through meetings; the WFP Azerbaijan report goes beyond this to a more detailed level of analysis that includes joint planning and implementation activities, as well as knowledge sharing.

As was found in 2002, reports tend to pay greater attention to interagency coordination; their relative lack of attention to coordination with government and local authorities may be part of the overall poor focus on capacity development in humanitarian action, noted in Chapter 3.

#### Evaluation of consideration given to protection (4.3iii)

Protection continues to be the cross-cutting area least well covered by EHA. Sixty-eight per cent of reports made no mention of protection, a similar finding to 2002, and 92 per cent were rated as unsatisfactory or poor, a worse performance than 2002 when the equivalent was 79 per cent. This despite a considerably greater number of reports on complex emergencies this year.

One of the reasons protection is so poorly covered is that evaluators may see protection as the exclusive mandate of the ICRC and UNHCR. Also, most evaluations have a fairly narrow sectoral focus and do not tend to look far beyond the 'technical' specifics of the intervention – such as the kinds of food provided and to whom, or how many pumps were sunk and whether they are still functioning. In terms of coverage, most NGO and ECHO reports did not cover this area. Conversely, it is interesting to note that of the WFP reports, and although WFP does not in general advocate a rights-based approach, the evaluation of the WFP Angola intervention (April 2002) integrates a detailed discussion of food-related protection issues, including the need for WFP to develop its programme to maximise protection; analysis of whether to refuse to distribute food in cases of forced displacement; and the potential for the provision of food aid decreasing the security of primary stakeholders. Individual evaluators attuned to protection questions can make recommendations on this issue even if this is not required by the TOR. But once again whether they do or not comes back to the question of how far the evaluator should advocate on controversial issues.

The other report included in the meta-evaluation that covered protection thoroughly is that of the Norwegian Ministry of Foreign Affairs (November 2001) – as might be expected in a report on a national Red Cross organisation. The UNHCR (May 2002) evaluation of its work on the protection of children, not

covered in the meta-evaluation because of its thematic focus, nevertheless provides an example of good practice and could be used by other agencies as an example as to what can be achieved in the evaluation of protection. This report clearly analyses the ways in which agency policies and principles on protection were applied. The methodology for this evaluation includes a strong focus on primary stakeholder consultation, and the evaluation notes that triangulation has been carried out – although the process by which this took place is not evident. While overall the evaluation is very rigorous, at times its conceptual discussion could have been complemented by more fieldbased observations; and because of limited focus on impact it is sometimes difficult to determine why conclusions were drawn, for example, concerning linking social and legal protection foci.

#### Evaluation of consideration given to gender equality (4.3iv)

Attention to gender equality was rated as less than satisfactory or poor in 81 per cent of cases, and in close to half the reports gender is not even mentioned.<sup>9</sup> Both the EU and NGOs performed badly. In contrast, six of the eight WFP reports were rated as good in this area and one as satisfactory; a considerable achievement given overall agency performance. In addition, WFP published a separate thematic report on its Commitments to Women. This impressive attention to gender equality has already been highlighted in Chapter 3 with reference to the results of WFP's interventions.

In most of the WFP reports attention to gender equality is mainstreamed throughout each report as well as being included in a separate section, the latter often being quite substantial. Also of note is an Annex in the reports which contains a checklist on 'Meeting the WFP Commitments to Women and Mainstreaming a Gender Perspective'. Each of the five commitments and their components are rated on a scale from very high to very low, and the reports include detailed narrative observations to complement the rating. The quality of reporting also suggests that WFP has made a commitment to hiring evaluators who have relevant skills in assessing gender equality. Overall, this is probably the most sustained attention to the evaluation of gender equality in EHA to date.

#### Consideration given to vulnerable/marginalised groups (4.3v)

In the QP definition, vulnerable and marginalised groups include the elderly, disabled, children, and people with HIV/AIDS. As many agencies have policies that

focus their programmes on the most vulnerable, one would expect substantial attention to these in the evaluation reports.

Despite this the picture is mixed, with only 36 per cent of reports rated as satisfactory or better. A problem identified in Annual Review 2002 – the failure of evaluators to disaggregate primary stakeholders – was also found this year, though to a lesser extent. There may be a hangover here from hiring evaluators who have technical expertise (e.g., water or health specialists) and not complementing this with social science expertise.

Of note is the attention given to HIV/AIDS in the ECHO reports on Burundi (ECHO, December 2001e, f, g). In general those reports that pay adequate attention to gender equality also tend to evidence a satisfactory level of attention to the vulnerable and marginalised, although only two reports were rated as 'good' in both these areas: the WFP reports on Iran and the Great Lakes (September 2002b, 2002). This suggests that a consistent attention to basic areas of social differentiation is hard for evaluators to achieve, and is an area where capacity development and training is needed.

## Assessment of Evaluation Practice

## M2.6

### Consultation with and participation by primary stakeholders (5.2i)

In order to achieve a satisfactory rating for this area, reports were required to provide adequate detail on the nature (e.g., focus groups) and scope (e.g., numbers by sex of those consulted) of consultation and participation. The failure to do this, noted in the two previous Annual Reviews, worsened this year. Only four evaluations were considered to have undertaken adequate consultation and describe in sufficient detail the consultation that occurred.

It is clear that many evaluators are talking to primary stakeholders, and some of the reports are peppered with their quotations or comments. Why evaluators do not detail the method behind these interviews adequately may be because:

- they do not see the relevance of including this information in the reports, thinking perhaps that it will lead to information overload;

**Table M8** Consultation and Participation During the Evaluation Process

| Area of Enquiry  | Rating         | % of Reports<br>Attaining<br>Rating (rounded) |
|--|----------------|---|
| <b>5.2i</b>  |                |   |
| Quality of consultation with and participation by primary stakeholders (beneficiaries and non-beneficiaries within the affected population) during the evaluation. | Good           | 5   |
|  | Satisfactory   | 5   |
|  | Unsatisfactory | 27  |
|  | Poor           | 63  |
| <b>5.2ii</b>   |                |   |
| Quality of consultation with, and participation by, other key stakeholders in the evaluation process.  | Good           | 6   |
|  | Satisfactory   | 8   |
|  | Unsatisfactory | 66  |
|  | Poor           | 20  |

- they are not aware of the importance of explicitly comparing the perspectives of different stakeholders to add credibility to the evaluation findings;
- much of the focus of EHA is on intra-institutional matters and field trips to project sites are rushed and given low priority.

However, consistent consultation with primary stakeholders, cross-referencing this with other perspectives and detailing the nature and scope of consultation will go a long way to overcoming one of the principal problems with EHA: its failure in many cases to establish credibility of evaluation methods. It will also help fulfil the participatory mandate of most agencies.

The exceptions this year prove that adequate consultation is possible. Of the four reports that were rated satisfactory or better (DEC, December 2001; WFP, April 2002; Oxfam, March 2002; Netherlands Ministry of Foreign Affairs, January 2002), the outstanding report was the DEC evaluation of British NGOs after the 2001 Gujarat earthquake, highlighted in Box M2. Extended consultation in the evaluation of the WFP intervention in Angola (WFP, April 2002) should also be noted.



**Box M2 Good Practice in Consultation with Primary Stakeholders**

DEC evaluations have consistently consulted with primary stakeholders (see Annual Review 2002). The evaluation of DEC agencies' performance in their response to the 2001 Gujarat earthquake (December 2001) is an excellent example of the levels of consultation that can be achieved:

- The evaluation notes the importance of attempting to empower communities through evaluation approaches that seek their active participation.
- The evaluation team included an Ahmedabad-based disasters institute, the Disaster Management Institute (DMI). DMI organised and conducted a survey covering 50 villages, and interviews with over 2,300 people. The inclusion of national researchers and consultants is a regular feature of DEC evaluations, unlike most other EHA.
- Interviews and focus groups were carried out using state-of-the-art participatory methodologies, and there was considerable attention paid to the location of consultation exercises in order to encourage the participation of as diverse a cross section of the community as possible.
- Specific attempts were made to include 'missing voices', including low status communities, the poorly educated, widows, women, the disabled and sick, those living on the outskirts of communities and working in nearby towns during the day. Timing and location of exercises and follow-up interviews attempted to include these groups.
- The methodology is detailed extensively.
- Quotes and comments from primary stakeholders are used effectively throughout the report to substantiate key points.

Use of the community survey has some weaknesses – for example, primary stakeholders were asked about the total intervention rather than specifically about the DEC agency intervention and this is not taken into account adequately in the analysis. Furthermore, conclusions in the report are sometimes at odds with the findings of the community survey. But it remains an impressive example of what can be accomplished given local expertise and the belief and willingness of the commissioning agency that such an exercise is worth pursuing.

### Consultation with and participation by other key stakeholders (5.2ii)

Although 86 per cent of reports were rated as unsatisfactory or poor in this area, much of this can be attributed to a requirement introduced to this year's QP that requires reports to explain the nature of such consultation (e.g., whether confidentiality was ensured). The rationale for this was that there can be a significant variation in responses dependent on the circumstances of an interview – for example, when a senior officer or other third person is present at the interview. As much of the weight in EHA rests on interviews with agency staff, it was thought important to include an assessment of how far the evaluation team facilitated independent expression of views.

The majority of evaluations did involve significant discussions with key stakeholders. However, they also featured generic problems. These included:

- consultation with only one set of stakeholders, usually agency staff, to the exclusion of national and local governments;
- lack of detail on the nature of the consultation, e.g., where it took place, who was present, or whether a questionnaire was used;
- failure to provide a list of key stakeholders consulted.

As with consultation with primary stakeholders the second and third bullet points could easily be corrected in many cases. Including adequate information on these areas will strengthen the credibility of reports.

### Quality of application of standard EHA criteria (5.3ii)

Application of the DAC/OECD criteria is one of the stronger areas of EHA. Reports rated highly in the evaluation of effectiveness, relevance/appropriateness, and sustainability/connectedness, where there is a range of good practice. This suggests that use of these criteria has been mainstreamed into EHA and to a lesser extent into evaluation of coverage. Efficiency, impact and, in particular, coherence, fared less well.

First, the good practice. Sixty-eight per cent of reports were rated as satisfactory or good in their evaluation of relevance/appropriateness. Reports rated as good include DEC (December 2001), WFP (December 2001, January 2002), and CARE

**Table M9** Application of Methods, Criteria and Standards

| Area of Enquiry   | Rating         | % of Reports<br>Attaining<br>Rating (rounded) |
|---|----------------|---|
| <b>5.3ii</b> Quality of Application of EHA<br>criteria in the assessment of intervention. |                |   |
| <b>a</b>  |                |   |
| Efficiency (including cost-effectiveness)   | Good           | 6   |
|   | Satisfactory   | 28  |
|   | Unsatisfactory | 46  |
|   | Poor           | 20  |
| <b>b</b>  |                |   |
| Effectiveness (including timeliness)  | Good           | 11  |
|   | Satisfactory   | 63  |
|   | Unsatisfactory | 23  |
|   | Poor           | 3   |
| <b>c</b>  |                |   |
| Impact  | Good           | 6   |
|   | Satisfactory   | 29  |
|   | Unsatisfactory | 41  |
|   | Poor           | 24  |
| <b>d</b>  |                |   |
| Relevance/appropriateness   | Good           | 17  |
|   | Satisfactory   | 51  |
|   | Unsatisfactory | 23  |
|   | Poor           | 9   |
| <b>e</b>  |                |   |
| Sustainability/connectedness  | Good           | 14  |
|   | Satisfactory   | 59  |
|   | Unsatisfactory | 22  |
|   | Poor           | 5   |
| <b>f</b>  |                |   |
| Coverage  | Good           | 21  |
|   | Satisfactory   | 35  |
|   | Unsatisfactory | 26  |
|   | Poor           | 18  |
| <b>g</b>  |                |   |
| Coherence   | Good           | 0   |
|   | Satisfactory   | 12  |
|   | Unsatisfactory | 15  |
|   | Poor           | 73  |

(September 2002). The report on WFP's interventions in Somalia, for example, analysed the overall strategy of delivery of food aid (included in the discussion of food aid in Chapter 3), and integrated considerable detail about the appropriateness of the ration.

Seventy-four per cent of reports were rated as satisfactory or better in their assessment of effectiveness and, in particular, as to whether inputs were turned into outputs (e.g., whether food aid was delivered or wells sunk). Seventy-three per cent were rated as satisfactory or better in assessment of sustainability/connectedness. This impressive rating suggests that evaluators are familiar with these concepts and for the most part have the ability to assess these areas, even though a significant minority of reports are under-performing. Only one report in the case of effectiveness, and two in the case of sustainability/connectedness, did not address these issues.

Problems with assessment of efficiency related mainly to a majority of reports not considering whether the intervention might have taken a less costly route to achieve its objectives – for example, whether different forms of procurement or logistics might have been more cost effective. However, as noted in Chapter 3, a minority of reports did cover areas such as differential costs between international and national staff, and local and international procurement.

The main problems with assessment of impact – where 65 per cent of reports were assessed as unsatisfactory or poor – was the inability of evaluators to look beyond the specific outputs of the intervention to the wider horizon and to examine any unintended consequences, whether positive or negative. One of the areas most often missed was consideration of how interventions were likely to affect socioeconomic relations over the longer term, including gender relations. However, this rating should be read in the context of general difficulties with the assessment of impact in the evaluation field, and it is usually acknowledged as one of the more difficult areas to evaluate.

'Coherence' is the least understood of the OECD/DAC criteria, and is often confused with 'coordination'. This is linked to the failure of evaluators to consider agency policy (see Section M2.4). Indeed the idea of considering whether a number of agencies' policies and strategic directions are similar was beyond the scope of most evaluations – perhaps because many looked at single agency interventions, and no good practice was identified.<sup>10</sup>

## Findings, Conclusions and Recommendations

## M2.7

## Overview

Evaluations were weak in terms of making proactive efforts to disseminate report findings, in particular to primary stakeholders, as well as in attempting to ensure that

**Table M10** Quality of Findings, Conclusions and Recommendations

| Area of Enquiry   | Rating         | % of Reports Attaining Rating (rounded) |
|---|----------------|---|
| <b>5.4i</b>   |                |   |
| Quality of the sharing of the evaluation findings.  | Good           | 0                                       |
| a. Preliminary findings should be discussed with key stakeholders, including primary stakeholders, as the evaluation progresses.  | Satisfactory   | 3                                       |
|   | Unsatisfactory | 38                                      |
|   | Poor           | 59                                      |
| <b>5.4i</b>   |                |   |
| Quality of the sharing of the evaluation findings.  | Good           | 0                                       |
| b. The draft evaluation report should be shared with key stakeholders, and feedback integrated into the final report or included as an Annex.   | Satisfactory   | 3                                       |
|   | Unsatisfactory | 13                                      |
|   | Poor           | 84                                      |
| <b>5.4ii</b>  |                |   |
| Quality of conclusions arising from findings.   | Good           | 17                                      |
|   | Satisfactory   | 45                                      |
|   | Unsatisfactory | 28                                      |
|   | Poor           | 10                                      |
| <b>5.4iii</b>   |                |   |
| Quality (including feasibility) of recommendations.   | Good           | 22                                      |
| a. Recommendations should respond to the main conclusions; reflect consultation with all key stakeholders; and understanding of the commissioning organisation; and potential constraints to follow-up. They should be clear, relevant and implementable with each ideally accompanied by implementing options. | Satisfactory   | 61                                      |
|   | Unsatisfactory | 33                                      |
|   | Poor           | 3                                       |
| <b>5.4iii</b>   |                |   |
| Quality (including feasibility) of recommendations.   | Good           | 0                                       |
| b. The evaluation report should suggest a prioritisation (e.g., into macro or structural, micro or easily achievable) and timeframe for follow-up and suggest where responsibility should lie if this is not indicated in the TOR.  | Satisfactory   | 6                                       |
|   | Unsatisfactory | 63                                      |
|   | Poor           | 31                                      |

recommendations were followed up and lessons learnt. They thus undermine the whole purpose of evaluation. When accepting evaluation contracts evaluators have a responsibility to attempt to improve agency performance. In general, more time needs to be spent working with agencies to establish what kinds of recommendations are feasible, what recommendations should be prioritised, and who should be responsible for follow-up. Recommendations are too often tacked on in long lists at the end of reports; worse, they are dispersed throughout the report and not included in the Executive Summary.

#### Sharing of findings (5.4ia & b)

A significant majority of reports include brief details as to feedback mechanisms, usually through end-of-mission meetings with agency country staff, workshops in-country, and debriefings in the HQ of the evaluated agency. There were also some, although fewer, details on circulation and feedback of report drafts. Both mechanisms are now often built into evaluations as a matter of course, but the reports are largely silent on how interaction during debriefings, and comments on drafts, affected the final conclusions and recommendations – even though it is well known that there is usually a period of bargaining between evaluators and commissioning agencies between draft and final versions of a report, particularly concerning phrasing and inclusion of unwelcome findings.

In this year's QP a requirement was added that evaluators needed to share preliminary findings with primary stakeholders (5.4ia) in order to be considered satisfactory. Only one evaluation managed this – the evaluation of Oxfam's intervention in Burundi (March 2002) where the evaluators presented evaluation results to stakeholders, including community hygiene and water committee members, local and national government representatives, and donor and INGO staff.

Part of the reason for the failure to take evaluation findings back to primary stakeholders for discussion and verification probably stems from the format of evaluation missions, which usually start and end in the national capital. There is also usually a major gulf between primary stakeholders and national capital-based staff of NGOs, donors and governments, which makes inviting primary stakeholders to national capital feedback meetings an option rarely considered. This means that what is generally considered good evaluation practice – verifying results with stakeholders – does not take place in EHA.

### Quality of conclusions (5.4ii)

Sixty-two per cent of reports rate as satisfactory or good in terms of conclusions flowing logically from, and reflecting, the report's central findings. The assessors mainly considered whether there was a clear connection between findings and conclusions and did not take into account the basis of findings, which is covered mainly under Section M2.2 above. If the latter area had been considered then the rating on conclusions would have been considerably lower.

### Quality of recommendations (5.4iiia & b)

This year two separate aspects were included in the attempt to differentiate particular strengths and weaknesses: clarity and quality. Writing recommendations is an art. It is perhaps even the most important part of evaluation practice and not something evaluators can be expected to do without training or guidance.<sup>11</sup> While some of the failure of uptake of recommendations is due to political factors within and between agencies – itself something evaluators should be aware of when writing recommendations – this also partly results because recommendations are inadequately crafted.

In 5.4iiia, the assessors mainly focused on whether recommendations were clearly written, relevant, and responded to the main conclusions. While this section of the QP also included a requirement that recommendations be implementable and demonstrate an understanding of the commissioning organisation, it was not possible for the assessors to judge this accurately given the wide range of countries and agencies involved.

Eighty-four per cent of reports were rated as satisfactory or better in terms of this first area, the highest rating of any QP Area of Enquiry. Recommendations tended to be clearly phrased and followed on from conclusions

However, in relation to 5.4iiib, no reports met the requirement of producing recommendations that were: a. prioritised; b. included a timeframe for follow-up; and c. suggested where responsibility for follow-up should lie. Many reports included long lists of recommendations, sometimes stretching to several pages and sometimes dispersed unhelpfully through the report.

In light of this, evaluators should consider taking a more proactive approach to the writing of recommendations:

- noting four or five recommendations they see as central in the Executive Summary section;
- providing a suggested timeframe for each of these recommendations;
- naming specific agency positions (e.g., project manager) responsible for follow-up; if that is not possible, a department or unit.

### Quality of Report Coverage, Legibility and Accessibility

M2.8

| Table M11 Report Coverage, Legibility and Accessibility |                |   |
|---|----------------|---|
| Area of Enquiry   | Rating         | % of Reports Attaining Rating (rounded) |
| <b>6.i</b>  |                |   |
| Quality of the coverage of the evaluation report.       | Good           | 23                                      |
|   | Satisfactory   | 35                                      |
|   | Unsatisfactory | 42                                      |
|   | Poor           | 0                                       |
| <b>6.ii</b>   |                |   |
| Quality of the format of the report.                    | Good           | 15                                      |
|   | Satisfactory   | 19                                      |
|   | Unsatisfactory | 7                                       |
|   | Poor           | 59                                      |
| <b>6.iii</b>  |                |   |
| Accessibility of the report.                            | Good           | 21                                      |
|   | Satisfactory   | 47                                      |
|   | Unsatisfactory | 29                                      |
|   | Poor           | 3                                       |
| <b>6.iv</b>   |                |   |
| Quality of the executive summary.                       | Good           | 21                                      |
|   | Satisfactory   | 46                                      |
|   | Unsatisfactory | 21                                      |
|   | Poor           | 12                                      |



### Quality of report coverage, legibility and accessibility (6.i & ii)

Note: the analysis of these two areas of enquiry only involves the 28 reports that included a TOR as coverage and format were rated against the requirements of the TOR.

With reference to the first area, evaluation reports were required to cover adequately all areas specified in the TOR in addition to any further factors likely to effect the performance of the intervention. The fact that in 58 per cent of cases where TOR were present there were areas missed by evaluators suggests that commissioning agencies are not using the TOR as a means of holding evaluators accountable to their agreement with the agency. Across the areas covered in the TOR, there was no one area that stood out as consistently missed by evaluators. However, five reports did not cover one or more of the DAC criteria, and three did not cover gender equality issues, all as required in the TOR.

In terms of report format, a majority of reports were rated as unsatisfactory or worse as the TOR did not provide a template format to follow (although in a number of cases, including the WFP reports, a required format is mentioned in the TOR but not included in the version received by ALNAP). Commissioning agencies not including a required format miss an important opportunity to provide evaluators with guidelines as to their priorities, as well as to promote greater attention to areas generally missed in EHA such as protection and gender equality. An exception were the ECHO reports which were found to be generally strong in both setting a required format and ensuring that this was followed.

### Accessibility of the report and Executive Summary (6.iii & iv)

Reports were generally well written in clear English or French, although more reports could have included visual aids such as maps, tables and diagrams. A number of reports included long stretches of unbroken text, trying for even the most patient and interested reader. The quality of Executive Summaries was satisfactory or better in 67 per cent of cases, and removing those four reports where no Executive Summary was included (a major oversight), performance in this area can be considered adequate. For the 21 per cent of reports rated as unsatisfactory, the main issue was failure to include all key report recommendations in the Executive Summary. There were no significant differences between ECHO, the UN and NGOs in these areas, except in the case of ECHO reports in English, which tended to lack clarity.

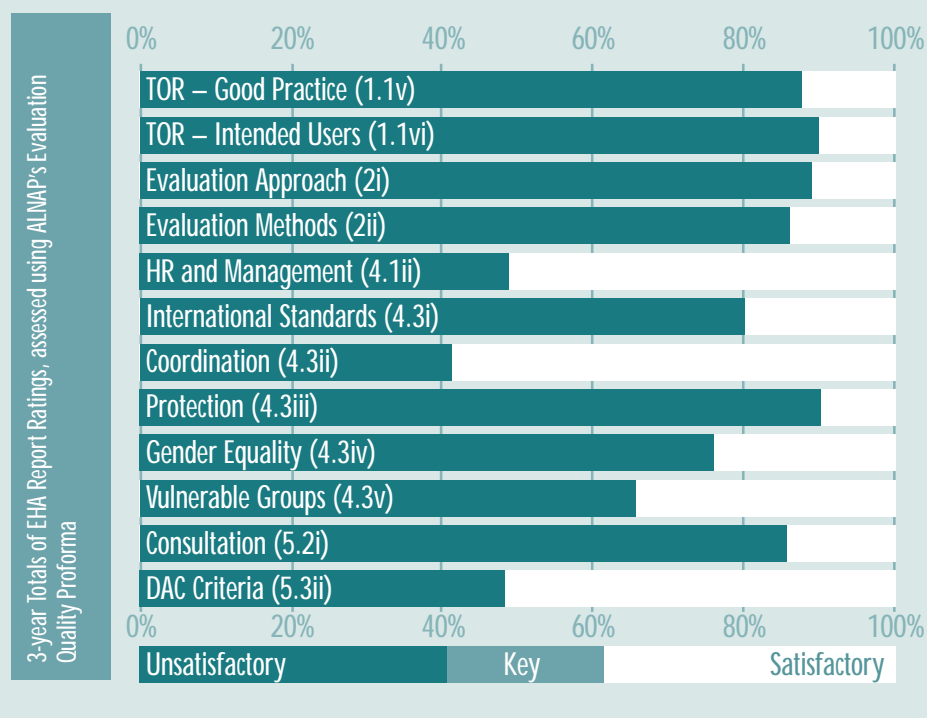
## Year-on-year Analysis of 127 Evaluations of Humanitarian Action (2000-2002)

M3

### Box M3 Year-on-year Analysis of 127 Evaluations of Humanitarian Action (2000-2002)

This year we report on some comparative areas that are key to successful evaluation over the period of the three Annual Reviews, based on the 127 reports assessed against the QP for 2000-2002. Methodological details related to this comparative analysis can be found in Section M1.6 above. The criteria against which reports were rated can be seen in the Guidance Notes section of the QP at the end of this section.

Twelve QP Areas of Enquiry are compared, as set out in the Figure below where aggregate results for the three year period are presented.



**Box M3** Year-on-year Analysis of 127 Evaluations  
of Humanitarian Action (2000–2002) *continued*

| TOR   | Satisfactory % | Unsatisfactory % |
|---|----------------|------------------|
| Quality of TOR statement on expectation of good practice in approach and method | 11             | 89               |
| Quality of TOR statement on intended use and users of evaluation outputs        | 8              | 92               |

Evaluation reports were found to be weak in both of these areas. In general, reports did not specify adequately the key methodological tools that evaluators should use. It was also rare for TOR to outline clearly the intended use of evaluation reports; failure to do this adds to the likelihood that the findings of these reports will not be fully used.

| Delineation of Methodology                         | Satisfactory % | Unsatisfactory % |
|--|----------------|------------------|
| Appropriateness of the overall evaluation approach | 10             | 90               |
| Appropriateness of the evaluation methods selected | 12             | 88               |

In terms of outlining, explaining and providing a rationale for the evaluation approach, performance was generally unsatisfactory, in particular in 2002 and 2003. EHA is atheoretical and as such derives little direction from wider evaluation thinking. For example, the debate over the relative emphasis to be placed on lesson learning and accountability in EHA has also been taking place in the wider evaluation field, but EHA practitioners have made few linkages. This is not to suggest that every evaluator needs to become a specialist in evaluation theory. Far from it. But commissioning agencies and evaluators do need to have a broad understanding of the strengths and weaknesses of different evaluation approaches so as to avoid common pitfalls and make EHA as rigorous as possible.

A small minority of reports achieved good practice in terms of delineating the methodology that was to be used. However, most reports note only basic details of the methodology, which in turn undermines the credibility of their findings.

**Box M3** Year-on-year Analysis of 127 Evaluations  
of Humanitarian Action (2000–2002) *continued*

|   | Satisfactory % | Unsatisfactory % |
|---|----------------|------------------|
| Quality of the Evaluation of Agency's Management and Human Resource Practices | 51             | 49               |

This is a strength in EHA, with over 50 per cent of evaluations rating as satisfactory or better each year. Evaluators have consistently examined issues such as staff turnover, HQ-field communication, and security. However this can be considered both a strength and a weakness because the focus on institutional issues may detract from other areas, such as consultation with primary stakeholders or international standards.

| Cross-cutting Themes                                   | Satisfactory % | Unsatisfactory % |
|--|----------------|------------------|
| Evaluation of use of international standards           | 20             | 80               |
| Evaluation of co-ordination                            | 58             | 42               |
| Evaluation of protection                               | 10             | 90               |
| Evaluation of gender equality                          | 26             | 73               |
| Evaluation of consideration to vulnerable/marginalised | 36             | 64               |

The cross-cutting theme that consistently scored well was coordination, which is related to the ability of evaluators to cover institutional factors. In the other four theme areas reports performed consistently poorly except in the case of consideration to the vulnerable and marginalised where performance was somewhat better. The link between international standards, protection and gender equality is that they deal with rights-based issues that are often controversial; these are the issues that are most often left out of evaluation TOR and with which evaluators appear to have the least skills. Protection is particularly poorly covered, with 92 per cent of the reports in 2002 and 79 per cent of reports in 2001 assessed as unsatisfactory or poor. The Red Cross/Red Crescent Code of Conduct and the Sphere standards are also not generally used.

This is a central gap in EHA, which is clearly a long way away from integrating a rights-based approach into a wider evaluative process. ALNAP can play an important role in terms of getting this issue on the agenda of commissioning agencies and evaluators.

### Box M3 Year-on-year Analysis of 127 Evaluations of Humanitarian Action (2000–2002) *continued*

|  | Satisfactory % | Unsatisfactory % |
|--|----------------|------------------|
| Quality of Consultation with and Participation by Primary Stakeholders | 13             | 87               |

This is a further area of weakness. Despite some good practice, EHA could rightfully be accused of systematically ignoring the views and perspectives of primary stakeholders in favour of those of institutional actors, particularly agency staff. This undermines its credibility and continues in the vein of treating primary stakeholders as passive recipients of aid rather than active participants in their own recovery. This agency-centric perspective will only change if commissioning agencies insist on adequate primary stakeholder consultation and participation. The constraints to this, particularly security issues, should not of course be underestimated. But an equally important constraint would appear to be the structure of evaluation missions which are usually short forays by foreign-based evaluators, with a focus on national capitals.

|                                 | Satisfactory % | Unsatisfactory % |
|---------------------------------|----------------|------------------|
| Application of the DAC Criteria | 50             | 50               |

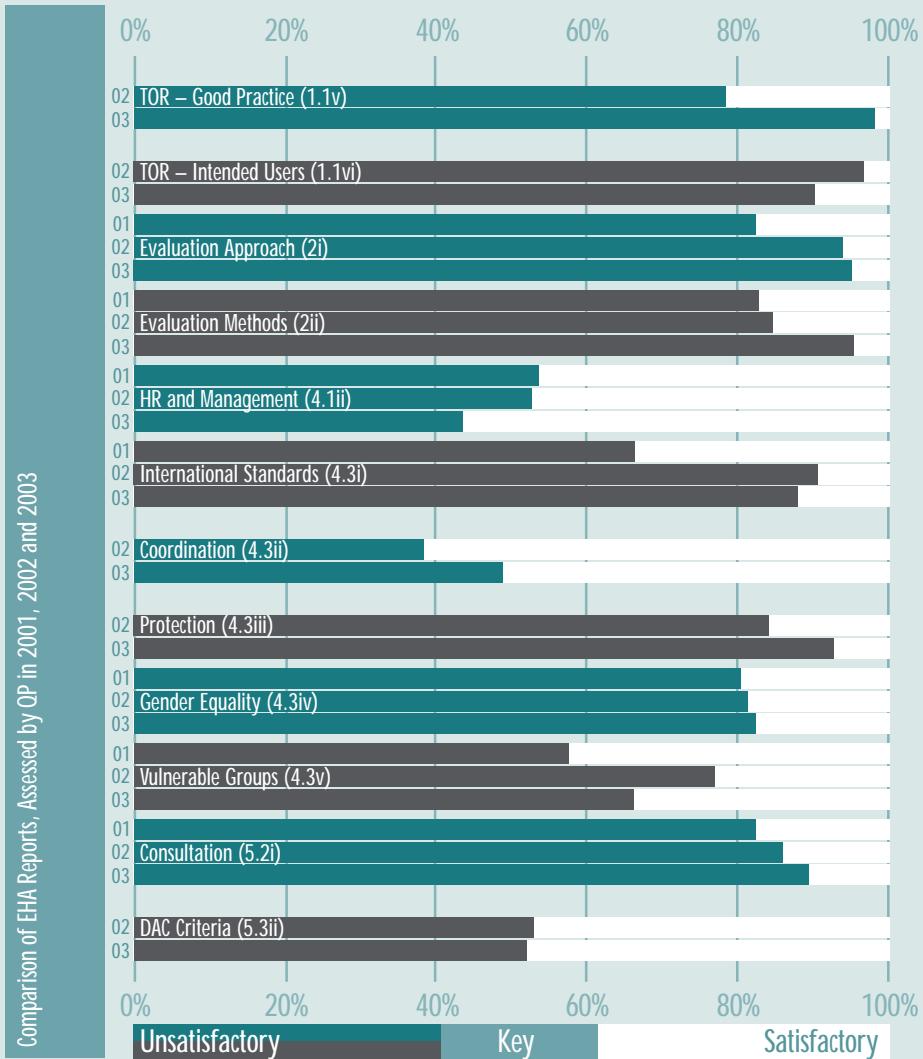
This analysis is based on reports assessed for this and last year only, as the QP for 2001 did not use comparable phrasing. Results for this year have been aggregated across the seven criteria. Much EHA is organised around the DAC criteria, as reflected in most evaluation TOR. Application of the DAC criteria is one of the stronger areas of EHA, with the third highest rating of the 12 areas covered in this Box. Overall it is possible to conclude that evaluators have had reasonable success with their application and that they have become EHA's central evaluative tool.

#### Year-on-Year Improvement?

As several of the areas of enquiry cover only this and last year, and because of changes in the Proforma over time, it is difficult to come to definitive conclusions concerning year-on-year improvement in EHA. The areas where there may have been some improvement are in the evaluation of management and human

**Box M3** Year-on-year Analysis of 127 Evaluations of Humanitarian Action (2000–2002) *continued*

resources as well as the mainstreaming of DAC criteria. In this latter area, however, only some of the criteria are being consistently used (see Section M2.6).



Given the emphasis on results-based planning in many agencies, one would expect to see ongoing improvement in EHA. Over the next two years it will be important to assess progress in key evaluation areas such as attention to international standards, gender equality, and consultation with and participation of primary stakeholders. Some suggestions on standards and target-setting for agencies are included in the conclusions to this meta-evaluation (Section M3).

### Interagency Differences?

Finally, were there any marked differences between the UN system, ECHO and NGOs as far as evaluation quality is concerned? All actors have their strengths and weaknesses, some of which can be highlighted as follows:

- ECHO and NGOs did relatively well in assessment of coordination, management and human resources and attention to the vulnerable, but relatively poorly in detailing evaluation processes and paying attention to adherence to international standards, protection and gender equality.
- Overall, UN agencies performed best in 10 of the 12 QP areas considered in this Box (only six areas are covered), often by a considerable margin. This relates in particular to strong evaluation performance by WFP and UNHCR. Even so, UN agencies failed to achieve a 50 per cent satisfactory rate in six of the areas considered.

## Conclusions

M4

This year's assessment of 39 reports revealed some improvement in evaluation performance, but also highlighted ongoing weaknesses. Good practice, this year in the case of WFP and the DEC, illustrates what is possible given resources, capacity and mindset. A common theme this year has been the need to understand and measure changes in social processes more thoroughly – in particular, power relations, gender relations and indigenous coping strategies. Commissioning agencies and evaluators should reflect on whether their evaluation's consideration of these areas is adequate.

Some of the key areas commissioning agencies and evaluators should pay attention to over the next year are as follows:

M

### Evaluation Focus

- Ensure that protection issues and reference to international standards are included in TOR, where relevant.
- Bring to the evaluator's attention relevant agency policies, including the gender equality policy if it exists, and ensure that the need to evaluate against agency policy is clearly set out in the TOR.
- Adequately evaluate primary stakeholder participation and consultation.
- In the context section of the report, note the past involvement of the agency in the affected area, any partnerships that have built up, and how these affected the intervention.
- Pay particular attention to the DAC criteria which may be less well covered in evaluations, in particular impact, efficiency, and coherence.
- Ensure that data in reports is disaggregated by socioeconomic status, ethnicity and sex.

### Evaluation Process

- Look for innovative ways to disseminate report findings, for example, through thematic summaries or key sheets. Follow-up informally with colleagues to see if recommendations have been followed. If they have not, analyse why.
- Promote primary stakeholder consultation and participation, and ensure that there is a requirement to do this in the TOR.
- Ensure that the methods used provide a credible basis for conclusions and that the description of the method fully reflects what the evaluation team has done – in particular in relation to consultation and participation of primary stakeholders, and the nature of participation of other key stakeholders.
- Note how confidentiality and dignity of respondents is ensured.
- Establish or build on contacts with evaluators from affected countries and consider including them in evaluation teams or making it a requirement in



tenders that at least one person from the affected country be included on the team.

- Publicise the cost of the evaluation in the TOR so as to allow a comparative analysis of evaluation costs and to ensure that adequate resources are being allocated for evaluation purposes.

Systemic problems in the quality of EHA have been identified by the three year comparative analysis in Box M3. This suggests that ALNAP member agencies should, among other initiatives, consider developing a set of standards for improving their evaluation practice in some of the weaker areas of EHA. While potentially controversial, it is this author's view that it is unlikely that there will be a significant improvement in evaluation practice over the next few years based on capacity development alone. This is because unsatisfactory practice is resulting not only from lack of capacity, but also because commissioning agencies are not consistently enforcing good practice requirements.

Agencies are in many cases already committed – through their policies and evaluation guidance – to covering adequately a number of the weaker areas in EHA, such as consultation with and participation of stakeholders, gender equality, use of good practice in methodology, and international standards. So target-setting to meet EHA standards would also be an accountability mechanism to ensure that agencies fulfil their commitments. Standards could be adapted from the QP, as have the lists above; if the idea of target-setting against these standards is adopted in principle the specifics would need to be discussed in the ALNAP forum by all ALNAP Full Members.

# Endnotes

## Chapter 1

- 1 For a fuller explanation of the differences between upward and downward accountability, see HPG (2002).

## Chapter 2

- 1 This chapter is based on an original draft by Hugh Goyder, amended and edited by Tony Beck. Additional information and research was provided by Silva Ferretti.
- 2 Utilisation-focused evaluation focuses on use of the evaluation for intended users – i.e., working with stakeholders to determine the use of the evaluation during the evaluation process. Empowerment evaluation focuses on using the evaluation process as a means of empowering stakeholders.
- 3 ALNAP members met included UNICEF, OCHA, USAID, ECHO, ICRC, SPHERE, UNHCR, WFP, WHO, OXFAM and SC(UK). In addition phone interviews were conducted with the BRCS, and MSF (Holland) and a number of independent consultants. The members of the Peer Review Group included Ian Christoplos (Independent); Rune Skinnebach (ICRC); Silva Ferretti (independent), Louis Sida (SCF-UK); Rebecca Scheurer (USAID); Eleanor Monbiot, (WVI); and William Carlos (Ireland Aid).
- 4 The understanding of concepts will of course vary according to cultural and linguistic context. An ECHO field worker with long experience of the former Soviet Union countries pointed out that the normal translation for the word ‘monitoring’ in Russian is ‘control’.
- 5 It should be noted, however, that a significant proportion of evaluations in this year’s data set were being carried out on ongoing interventions, possibly because of

their extended length (see the Meta-Evaluation Chapter for more details).

- 6 The Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief also provides a useful framework for monitoring at a more philosophical and ethical level. The UK Disaster Emergency Committee’s *Gujarat Earthquake Evaluation* (DEC, 2002) was based on an assessment of the DEC agencies’ response against the Red Cross Code of Conduct.
- 7 As the MSF-Holland (1999) Monitoring Manual states: ‘monitoring includes all forms of communication: verbal and written, formal and informal, creating the potential for cross-checking information.’ There is a need for more detailed comparative research into the extent to which managers use the information generated by formal monitoring systems.
- 8 OCHA has developed a set of Operational Principles for Humanitarian Information Management and Exchange, including accessibility, inclusiveness, inter-operability, accountability, verifiability, relevance, objectivity, humanity, timeliness, and sustainability (OCHA, 2002).

## Chapter 3

- 1 Feedback from ALNAP Full Members on *Annual Review 2002* suggested that this format was useful and it has therefore been repeated. Food aid is probably the largest relief/rehabilitation sector in terms of funding; for 2000 and 2001 approximately 60 per cent of all global contributions to the CAP consisted of support to the food sector (OCHA, April 2002).
- 2 For example, WFP’s policy *From Crisis to Recovery*, where food assistance is provided to support longer term development in protracted emergencies, considered in the section on food assistance. ECHO’s policy states (ECHO, 1999:12): ‘[H]umanitarian

aid will seek where possible to bear in mind and remain compatible with longer-term developmental objectives.’

- 3 ALNAP’s definition of an external evaluation of humanitarian action is as follows: ‘A systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability, which is: commissioned by or in co-operation with the organisation(s) whose performance is being evaluated; undertaken either by a team of non-employees (external) or by a mixed team of non-employees (external) and employees (internal) from the commissioning organisation and/or the organisation being evaluated; assesses policy and/or practice against recognised criteria: efficiency, effectiveness/timeliness/coordination, impact, connectedness, relevance/ appropriateness, coverage, coherence and, as appropriate, protection; articulates findings, draws conclusions and makes recommendations.’
- 4 This chapter has been written by the same author in each of the three *Annual Reviews*, which has facilitated making comparisons. For this year, five French individual reports and one French synthesis report were included and summarised by Sylvie Robert, consultant to ALNAP. Comparative figures for *Annual Review 2001* were 49 individual evaluation reports and five synthesis reports; and for *Annual Review 2002*, 46 individual reports and nine synthesis reports. The number of reports is therefore fairly consistent over the three years, although the source of reports and sectoral and geographical focus has varied each year.
- 5 Thanks are due to two peer reviewers, Susanne Jaspers and Jeremy Shoham, for comments on this section.
- 6 WFP is synthesising the findings of the PRRO evaluations, but the draft of this report was available only after this chapter went to press. However, the PRRO evaluations were discussed in a meeting with the WFP Evaluation Office staff in Rome in January 2003.
- 7 The *Annual Review 2002* (pp. 101–102) pointed out confusion in phrasing of objectives of WFP’s interventions as to whether they were maintaining or improving nutritional status. This confusion remains. For example WFP (April 2002: 9) has as one of its stated objectives to: ‘Maintain and improve the nutritional status of groups identified as nutritionally at risk ...’ See also the objectives set out in the terms of reference for WFP (January 2002) and WFP (December 2001a). This issue should be discussed at the Project Review Committee in WFP.
- 8 In his peer review of this chapter, Jeremy Shoham pointed out that the nutritional status of vulnerable populations may be worse than (some of) the evaluations conclude, drawing on recent evidence from *Field Exchange* and the *Report of the Nutritional Situation of Refugees and Displaced Populations*. Unfortunately it was not possible within the scope of this chapter – the main purpose of which is to summarise evaluation findings – to make direct comparisons between nutritional surveys reported in these publications and the evaluation reports. Noteworthy, however, is that evaluations have made no use of the *Report of the Nutritional Situation of Refugees and Displaced Populations* publications, despite the extensive nutrition information provided therein. The lack of credibility of some aspects of the evaluations reviewed this year is discussed in the meta-evaluation section.
- 9 This is also addressed in the WFP (1999) *Enabling Development*, but only in the

- context of development interventions.
- 10** Sen's entitlement theory has been followed by extensive debate which it is not possible to cover here, but see on its perceived failure to account for political economy, de Waal (1997).
- 11** It is unclear from this quote whether the recommendation is to provide blanket supplementary feeding or a general ration, but presumably the latter is meant. See also Oxfam (Bangladesh) which recommends targeted coverage for vulnerable groups such as pregnant women, and blanket coverage for everyone else on equity grounds. These findings can be contrasted to the DEC (December 2001:15, 24) Gujarat evaluation which found: 'People felt that organisations should distribute to the poor first, and not on a first-come, first-served basis. They wanted outsiders to ask communities who were the poor rather than rely on the views of leaders ... a particularly stark finding is that communities felt that relief was not given according to need.'
- 12** A recent Shelter Project draft report (2002:3) noted two contradictory findings in relation to what it terms transitional shelter (or shelter which provides covered living space in the interim period between primary stakeholders being forced to leave their home and being able to return or move to permanent new shelter): '[F]irstly, a number of case studies demonstrated high quality, appropriate responses; and secondly, evaluations of large emergencies were consistently critical of [the] transitional settlement sector specifically. This may indicate high quality field staff operating with insufficient policy, guidelines and organisational support, within poor coordinating structures. Conversely, it may indicate that evaluators had few tools and examples of best and worst practice with which to assess whether programmes were appropriate and effective.' <shelterproject.org> is currently working to develop policies and guidelines for the transitional shelter sector.
- 13** This is discussed in more detail in *Annual Review 2002*.
- 14** CARE (March 2002); DEC (December 2001); WFP (September 2002b, April 2002, September 2002); Norway Ministry of Foreign Affairs (November 2001); Oxfam (March 2002, May 2002, June 2002); ECHO (December 2001c, December 2001g); WHO (December 2002); UNHCR (December 2001); and IFRC (September 2002).
- 15** As Minear (2002) points out, terms such as 'international' and 'national' carry connotations related to expertise. These terms are used here as the best alternative, recognising the bias they may introduce.
- 16** Further details and references can be found in Mayhew (2002), which deals mainly with international staff.
- 17** Conceived around five country case studies (Afghanistan, Angola, Colombia, eastern DRC and Sri Lanka), the Global Study was commissioned by ALNAP to: assess current consultation and participation practice in a range of emergency contexts; identify examples of good practice; identify gaps or inadequacies in current practice and contributing factors; and improve understanding of participation and consultation practice. The study seeks wherever possible to promote successful mechanisms and initiatives identified, with the overarching objective of providing practical guidance for humanitarian agencies and their personnel. Outputs include: individual monographs linked to each of the case studies, a Practitioner Handbook (the core output of the study with front line practitioners as its primary audience), and an Overview Book in

which the issues arising from the studies and an extensive literature review are discussed in greater detail. Copies can be obtained from the ALNAP Secretariat ([www.alnap@odi.org](mailto:www.alnap@odi.org)).

- 18 Blue Flag volunteers are community members involved in health education programmes.
- 19 At the 2001 G8 Summit, leaders endorsed an agreement reached by the OECD-DAC. This agreement commits donors to untie aid on seven categories of aid to least developed countries: balance of payments and structural adjustment support; debt forgiveness; sector and multi-sector program assistance; investment project aid; import and commodity support; commercial services contracts; and ODA to non-governmental organisations for procurement-related activities. The agreement does not apply to food aid, all free-standing technical cooperation and management services arrangements (CIDA, 2002).
- 20 The Joint Evaluation of Emergency Assistance to Rwanda (JEEAR, 1996) found that (Study 3:101–2): ‘[T]here is ample evidence that the airlift, or at least substantial parts of it, continued for several weeks longer than was required ... It is perfectly conceivable that much of the cargo airlifted from Entebbe to Goma and Kigali could have been tricked instead, potentially saving several million dollars.’
- 21 IFRC (2001) reports on a similar finding from floods in Viet Nam.
- 22 Results-based planning refers to planning that supports the systematised measurement of results.
- 23 For example, CARE (March 2002); Tearfund (May 2002); Norway Ministry of Foreign Affairs (November 2001); WFP (September 2002b, April 2002, September 2002, December 2001, January 2002,

December 2001a); Oxfam (May 2002, July 2002); UNHCR (December 2001); and FAO (September 2002).

## Chapter 4

- 1 Staff may of course move from agency to agency within the sector, thus retaining capacity.
- 2 Some ALNAP Full Members have pointed out that three years is too short a period to expect to see substantial improvements in evaluation practice.
- 3 Taking into account ongoing work of other organisations, e.g. the Humanitarian Accountability Project.
- 4 On process monitoring, see Mosse et al (1998). The development of such tools is part of a wider movement in the evaluation field experimenting with participatory approaches to improve understanding of social process (e.g. Fetterman, 2002).
- 5 We have used the OECD/DAC (1999:22) guide as the best working definition of impact: ‘Impact looks at the wider effects of the project – social, economic, technical, environmental – on individuals, gender and age-groups, communities, and institutions. Impacts can be immediate and long-range, intended and unintended, positive and negative, macro (sector) and micro (household). Impact studies ask the question: what real difference has the activity made to the beneficiaries? How many have been affected?’

## Meta-evaluation

- 1 The ALNAP Quality Proforma (QP) will be sent for review and inclusion on the website of the Evaluation Centre of Western Michigan University, which has been running a project on checklists and maintains a website that currently includes 26 checklists as well as guidance on the development of checklists.

- 2 Peter Wiles has had extensive experience in the evaluation of humanitarian action, recently leading the DEC Kosovo evaluation and evaluations of the DFID/IFRC and DFID/ICRC institutional partnerships. Sylvie Robert also has significant experience of EHA in a number of countries including the Great Lakes region and the Balkans. Using one rather than two assessors for the reports in French means that there is greater chance of bias.
- 3 Rating was considered similar when the two assessors gave the same rating, or were one rating apart in the cases of 'A' (good) as opposed to 'B' (satisfactory), and 'C' (unsatisfactory) as opposed to 'D' (poor); and dissimilar in all other cases. A similar overall level of agreement was achieved for Annual Review 2002. See the QP at the end of this section for further information on the rating scale.
- 4 Thirteen evaluations were provided by bilaterals in 2000, five in 2001, and three in 2002. The 2001 set was skewed by eight reports from a DANIDA thematic evaluation. ICRC reports may be subject to greater confidentiality issues than other EHA.
- 5 A programme has been taken to mean a number of projects combined in a coherent fashion; a project has been taken to mean an individual intervention.
- 6 Agency standards tend to note that about 5 per cent of intervention expenditure should go towards monitoring and evaluation. Of this, agencies will need to determine what amount goes to evaluation. UNHCR has apparently developed strategic plans to commit 0.5 per cent of its total annual budget to evaluation activities by 2004 (ECHO, December 2001h). UNHCR (May, 2002) notes the cost of the evaluation, but only in an Annex supplied by the evaluation team.
- 7 The Standing Offer process operated by CIDA whereby consultants compete to provide services to CIDA in specific areas, including performance review, and are assessed against standard criteria before being accepted to the roster, could be developed for this purpose.
- 8 The DAC criteria are listed with summarised definitions in Annex 1 to the QP included at the end of this section.
- 9 These figures may be skewed by the requirement for this year that reports specifically assess agencies' performance against their gender policies (or note the lack of these) to achieve a satisfactory rating. While reference was made of the agency's policy on gender equality in the QP used for the previous two years, this was not specifically used to determine the cut-off point for a satisfactory rating.
- 10 In the summer of 2003 ALNAP will be producing a 'Guidance Booklet' to using the DAC/OECD criteria, which it is anticipated will help evaluators overcome constraints they are facing.
- 11 Patton (1997) makes the point from the utilisation-focused evaluation perspective that recommendations should be discussed and agreed upon by key stakeholders even before the evaluation report is written.

## Annex 1 Acronyms

|  |   |
|--|---|
| <b>AAR</b> After-Action Review   | <b>HQ</b> Head Quarters   |
| <b>ACF</b> Association of Charitable Foundations   | <b>IASC</b> Inter-Agency Standing Committee                                     |
| <b>ALNAP</b> Active Learning Network for<br>Accountability and Performance in<br>Humanitarian Action | <b>ICC</b> Interim Care Centres   |
| <b>ALPS</b> Accountability Learning and Planning<br>System   | <b>ICRC</b> International Committee of the Red<br>Cross                         |
| <b>BGVS</b> Bharat Gyan Vigyan Samiti  | <b>IDP</b> Internally Displaced Person  |
| <b>BRCS</b> British Red Cross Society  | <b>IFRC</b> International Federation of Red Cross<br>and Red Crescent Societies |
| <b>CAP</b> Consolidated Appeals Process  | <b>IMSMA</b> Information Management System for<br>Mine Action                   |
| <b>CCs</b> Collective Centres  | <b>INGO</b> International Non-governmental<br>Organisation                      |
| <b>CIDA</b> Canadian International Development<br>Agency   | <b>IRC</b> International Rescue Committee                                       |
| <b>DAC</b> Development Assistance Committee  | <b>JEEAR</b> Joint Evaluation of Emergency<br>Assistance to Rwanda              |
| <b>DANIDA</b> Danish International Development<br>Assistance   | <b>JEFAP</b> Joint Emergency Food Aid Programme                                 |
| <b>DEC</b> Disasters Emergency Committee   | <b>LRRD</b> Linking Relief, Rehabilitation and<br>Development                   |
| <b>DFID</b> Department for International<br>Development (UK)   | <b>LSO</b> Learning Support Office  |
| <b>DMI</b> Disaster Mitigation Institute   | <b>M &amp; E</b> Monitoring and Evaluation                                      |
| <b> DPRK</b> Democratic People's Republic of Korea   | <b>MAC</b> Mine Action Centre   |
| <b>DRC</b> Danish Refugee Council  | <b>MACC</b> Mine Action Coordination Centre                                     |
| <b>ECHO</b> European Commission Humanitarian<br>Aid Office   | <b>MSF</b> Medecins sans Frontieres   |
| <b>ECU</b> Emergency Coordination Unit   | <b>MSF-H</b> Medecins sans Frontieres - Holland                                 |
| <b>EHA</b> Evaluation of Humanitarian Action   | <b>NGO</b> Non-governmental Organisation  |
| <b>ERD</b> ALNAP Evaluative Reports Database   | <b>O &amp; M</b> Operations and Management                                      |
| <b>EU</b> European Union   | <b>ODI</b> Overseas Development Institute                                       |
| <b>EWS</b> Early Warning Systems   | <b>OECD</b> Organisation for Economic<br>Cooperation and Development            |
| <b>FALU</b> Food Aid Liaison Unit  | <b>OCHA</b> (UN Office) Coordination of<br>Humanitarian Assistance – see UNCHA  |
| <b>FAO</b> Food and Agriculture Organisation (UN)  | <b>OFDA</b> Office of Foreign Disaster Assistance<br>(USAID)                    |
| <b>FFW</b> Food For Work   | <b>PIA</b> People In Aid  |
| <b>Groupe URD</b> Groupe Urgence-Réhabilitation-<br>Développement                                    | <b>PRA</b> Participatory Rural Assessment                                       |
| <b>GTZ</b> Gesellschaft für Technische<br>Zusammenarbeit   | <b>PRRO</b> Protracted relief and recovery<br>operation                         |
| <b>HAP</b> Humanitarian Accountability Project   | <b>QP</b> Quality Proforma  |
| <b>HEP</b> Humanitarian Evacuation Programme   | <b>RBM</b> Results based management   |
| <b>HIC</b> Humanitarian Information Centres  | <b>R&amp;R</b> Rest and Relaxation  |
| <b>HPG</b> Humanitarian Policy Group (ODI)   | <b>RTE</b> Real-Time Evaluation   |
| <b>HPN</b> Humanitarian Practice Network (ODI)   |   |

- RRC** Rwandan Red Cross
- RRN** Relief and Rehabilitation Network
- SCF** Save the Children Fund
- SF** Supplementary Feeding
- Sida** Swedish International Development Agency
- TOR** Terms of reference
- UN** United Nations
- UNDP** United Nations Development Programme
- UNHCHR** United Nations High Commissioner for Human Rights
- UNHCR** United Nations High Commission for Refugees
- UNICEF** United Nations Children's Fund
- UNMAS** United Nations Mines Action Service
- UNOCHA** United Nations Office for the Coordination of Humanitarian Assistance
- UNOPS** United Nations Office for Project Services
- USAID** United States Agency for International Development
- WFP** World Food Programme (UN)
- WHO** World Health Organisation (UN)
- WVE** World Vision Ethiopia



## Annex 2 Glossary

- Accountability** ‘Accountability is the means by which individuals and organisations report to a recognised authority, or authorities, and are held responsible for their actions,’ (Edwards and Hulme, 1995).
- Agency mandate** An agency’s mandate is a formal statement approved by its governance mechanisms that articulates the *raison d’être* of its existence and the focus of its activity. In the case of inter-governmental and multilateral organisations their mandates derive from the international agreements that established them.
- Beneficiaries** The individuals, groups or organisations that benefit, directly or indirectly, from the assistance or services provided by others.
- Communities of practice** Groups of individuals sharing a common working practice even though not part of a formally constituted work team.
- Complex political emergency** A situation with complex social, political and economic origins which involves the breakdown of state structures, the disputed legitimacy of host authorities, the abuse of human rights and possibly armed conflict, that creates humanitarian needs. The term is generally used to differentiate humanitarian needs arising from conflict and instability from those that arise from natural disasters.
- DAC Evaluation criteria** (from OECD, 1999, pp 30-32):
- Efficiency** ‘measures the outputs – qualitative and quantitative – in relation to the inputs. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been used.’
- Cost-effectiveness** ‘... 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.’
- Effectiveness** ‘measures the extent to which the activity achieves its purpose, or whether this can be expected to happen on the basis of the outputs. Implicit within the criteria of effectiveness is timeliness (for if the delivery of food assistance is significantly delayed the nutritional status of the target population will decline).’
- Impact** ‘looks at the wider effects of the project – social, economic, technical, environmental – on individuals, gender, age-groups, communities, and institutions.’
- Relevance** ‘is concerned with assessing whether the project is in line with local needs and priorities (as well as donor policy) ... refers to the overall goal and purpose of a programme.’
- Appropriateness** ‘the need “to tailor humanitarian activities to local needs, increasing ownership, accountability, and cost-effectiveness accordingly” (Minear, 1994) ... is more focused on the activities and inputs.’
- Sustainability** ‘is concerned with measuring whether an activity or an impact is likely to continue after donor funding has been withdrawn ... many humanitarian 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. Larry Minear has referred to this as Connectedness, the need “to assure that activities of a short-term emergency nature are carried out in a context which takes longer-term and inter-connected problems into account” (Minear, 1994).’
- Coverage** ‘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”.

**Coherence** ‘refers to policy coherence, and the need to assess security, developmental, trade and military policies as well as humanitarian policies, to ensure that there is consistency and, in particular, that all policies take into account humanitarian and human rights considerations.’

**Downward Accountability** refers to ‘partners, beneficiaries, staff and supporters’ as opposed to Upwards accountability who are accountable to ‘trustees, donors and host governments’ (Edward & Hulme, 1995, pg 9)

**Disaggregation** Separation of data into its component parts.

**Evaluation** ‘The process of determining the merit, worth or value of something or the product of that process,’ (Scriven, 1991, p139).

**Evaluation of humanitarian action** A systematic and impartial examination of humanitarian action intended to draw lessons to improve policy and practice and enhance accountability. (ALNAP Annual Review, 2002, pg 201):

**Evaluation synthesis** An analysis of a series of evaluations to form an overall picture and assessment of the projects, programmes, policies or organisations that have been evaluated. Very similar to the term meta-analysis which is a particular approach to synthesising the results of studies of a common topic.

**Evaluative report** A report which, in some way, assesses and comments upon the performance of a project, programme, policy or an organisation or organisations. An evaluation report is a specialised type of evaluative report that is distinguished by its adherence to systematic evaluation procedures and recognised evaluation criteria.

**Humanitarian** ‘Being concerned with the

condition of man [sic] considered solely as a human being, regardless of his value as a military, political, professional or other unit,’ (Pictet, 1958, p96).

**Humanitarian action** Assistance, protection and advocacy actions undertaken on an impartial basis in response to human needs resulting from complex political emergencies and natural hazards.

**Humanitarian principles** A framework of principles derived from International Humanitarian Law which is ‘intended to guide and position humanitarian agencies ... in assisting and protecting those outside the limits of war in ways that are both ethical and practical,’ (Leader, 2000) that also places obligations onto parties to the conflict and attempts to regulate the conduct of a conflict.

**Humanitarian system** The group of organisations involved in the provision of humanitarian assistance and protection.

**Impartiality** An approach to the provision of humanitarian assistance and services which is non-discriminatory, proportionate to needs and free of subjective distinction. A guiding principle of organisations claiming to be humanitarian.

**Informating** Using IT not only to store but also to empower front-line staff to act on own initiative.

**Knowledge management** ‘Knowledge management is the systematic process of identifying, capturing, and transferring information and knowledge people can use to create ... and improve,’ American Productivity and Quality Center ([www.apqc.org/km](http://www.apqc.org/km)).

**Lesson-learning study** A study initiated by an organisation with the explicit objective of lesson-learning within that organisation, but that falls outside the full evaluation definition. A process that may be facilitated by external consultants but is generally an internal process.

**Logframe** Logical framework. Management

tool used to improve the design of interventions, most often at the project level. It involves identifying strategic elements (inputs, outputs, purpose, goal) and their causal relationships, and the assumptions or risks that may influence success and failure. It thus facilitates planning, execution and evaluation of an intervention.

**Meta-evaluation** The systematic evaluation of evaluations in order to determine their quality and adherence to established good practice in evaluation.

**Military humanitarianism** The provision of assistance and protection by military forces in response to humanitarian needs. A much-disputed term due to the difficulty, if not inability, of military forces to provide assistance on a genuinely non-discriminatory basis.

**Monitoring** (a working definition) is 'The systematic and continuous assessment of the progress of a piece of work over time ... It is a basic and universal management tool for identifying the strengths and weaknesses in a programme. Its purpose is to help all the people involved make appropriate and timely decisions that will improve the quality of the work.' (Gosling and Edwards, 1995 p. 81).

**Neutrality** Refers to the principle that to enjoy the confidence of all, an organisation may not take sides in hostilities or engage at any time in controversies of a political, racial, religious or ideological nature.

**Outcome-oriented evaluation** An evaluation by an independent team using investigative approaches emanating from accountability-oriented disciplines.

**Participatory evaluation** An evaluation process in which stakeholders play an active role in the design of the process, in its undertaking and in the development of the findings, conclusions and recommendations.

**Protection** 'Activities aimed at obtaining full respect for the rights of the individual in accordance with the letter and the spirit of the relevant bodies of law (i.e., human rights, humanitarian and refugee law) ... [which are] conduct[ed] impartially and not on the basis of race, national or ethnic origin, language or gender' (Von Flüe and De Maio, 1999).

**Results-based planning** The systematised measurement of results.

**Red Cross/Red Crescent NGO Code of Conduct** Principles of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Response Programmes. [www.ifrc.org/publicat/conduct](http://www.ifrc.org/publicat/conduct)

**Security** 'The protection of aid personnel and aid agency assets from violence' (Van Brabant, 2000).

**Social learning** 'The processes and products of people working together to discover and understand actual and preferred practice, about, in the case of humanitarian programmes, the best ways of meeting immediate survival needs in times of severe stress' (Apthorpe and Atkinson, 1999).

**Stakeholders** All those – from agencies to individuals – who have a direct or indirect interest in the humanitarian intervention, or who affect or are affected by the implementation and outcome of it.

**Triangulation** 'The use of several sources of information, methods or analysis to verify information and substantiate an assessment' (OECD, 2000, p16).

**Terms of reference** Terms of reference define the requirements and parameters for conducting an evaluation.

**Utilisation-focused evaluation** Utilisation-focused evaluation prioritises the needs of the primary intended users of the evaluation, in terms of the information they need for programme improvement and decision-making.

## Annex 3 References for Source Evaluations

by Commissioning Agency  
with Key Descriptors

**Objective (O)** Purpose of the evaluation as stated in the terms of reference, or general descriptor of the evaluation.

**Focus (F)** Main sectors/issues covered in the evaluation.

**Criteria (C)** Evaluation criteria used to assess the humanitarian action being evaluated.

**Comments (A)** Any additional information not covered above.

### Core Sample

#### Catholic Agency for Overseas Development (CAFOD)

■ CAFOD (July 2001) *After the Rains – a Greener Future: Environmental Evaluation of CAFOD's Response to the Flooding in Mozambique 2000* (London, UK: CAFOD)

Andersson, A. and I. Porsché

**O** Provide indications of the impact of CAFOD-supported relief and rehabilitation flood responses in Mozambique on the environment.

**F** Environment, shelter, water, sanitation, agriculture.

**C** Impact.

#### CARE International (CARE)

■ CARE International (March 2002) *Bosnia and Herzegovina: Quick Impact Facility* (London, UK: CARE)

Smith, D., Kirolos, T. and A. Wosner  
**O** Assess the impact of the 'Quick Impact' economic regeneration programme, from November 1999 to August 2002; and review the process of its implementation.

**F** Economic regeneration, refugee repatriation, livelihood indicators, socio-economic impact.

**C** Efficiency, impact, sustainability.

■ CARE (September 2002) *Evaluation of CARE Afghanistan's Emergency Response* (Atlanta, USA: CARE)

Cosgrave, J. and J. Baker

**O** Conduct an assessment and make recommendations concerning the CARE International response to the Afghan crisis, from September 2001 to May 2002, according to specified criteria.

**F** Food distribution, water reconstruction, shelter, rehabilitation.

**C** Appropriateness, efficiency, impact, coverage, connectedness, coordination.

#### Caritas Internationalis (Caritas)

■ Caritas (January 2002) *Five-year Review of the CARITAS Programme in the Democratic People's Republic of Korea (DPRK)* (Hong Kong SAR, PR China: Caritas Hong Kong)

Smith, H.

**O** Document and describe the work of the CARITAS programme over the 5 years up to 2001; examine the appropriateness of the support and identify learning points.

**F** Food aid, food security, agriculture, monitoring.

**C** Appropriateness, coverage, efficiency, impact, connectedness, coherence.

#### Danish Refugee Council (DRC)

■ DRC (April 2002) *Synthesis Report: Review of IGA Programmes of Danish Refugee Council* (Copenhagen, Denmark: DRC) Quest Consult De Klerk, T.

**O** Assess how the income generation projects (IGA) work as strategies, and their relevance within the context of local and national priorities, their role in post-conflict rehabilitation, best practices and lessons learned, based on field studies undertaken between November 2001 and February 2002.

**F** Capacity building, rehabilitation, repatriation, income generation, food security

**C** Appropriateness, sustainability, impact, effectiveness, efficiency.

**A** Field studies of approximately one week's duration were undertaken in Azerbaijan, Georgia, Serbia, Montenegro and Somaliland.

### Department for International Development – UK (DFID)

■ DFID (January 2002) *Evaluation of the Strategic Partnership between DFID and the International Federation of Red Cross and Red Crescent Societies* (London, UK: DFID) Valid International Wiles, P., Sandison, P., Porter, T. and M. Reddick

**O** Assess the impact and relevance of the strategic partnership between DFID and the Red Cross/Crescent Federation, over the 3 year period since 1999.

**F** Strategic partnership, organisational capacity building, disaster preparedness.

**C** Relevance, impact, connectedness, effectiveness, efficiency.

### Disasters Emergency Committee (DEC)

■ DEC (December 2001) *Independent Evaluation of Expenditure of DEC India Earthquake Appeal Funds: Volume One – Executive Summary; Volume Two – Full Evaluation Report; Volume Three – Sector Reports* (London, UK: DEC) Humanitarian Initiatives/Disaster Mitigation Institute/Mango

**O** Evaluate the DEC appeal in response to the India earthquake of January 2001 for the purposes of accountability, lesson learning to improve future performance and for compliance with the Red Cross Code of Conduct.

**F** Earthquakes, targeting, shelter, reconstruction, rehabilitation, coordination, beneficiaries, international codes and standards, financial management.

**C** Timeliness, appropriateness, cost-effectiveness, impact, coverage, connectedness, coherence.

### European Community Humanitarian Office (ECHO)

■ ECHO (October 2001) *Evaluation of ECHO's Global Plans 2000 and 2001 – Democratic Republic of Congo: Synthesis report* (Aachen, Germany: ECHO) Germax Gerli GmbH Kunze, M., Barthes, O., Logez, M. and F. von Massow

**O** Assess appropriateness and effectiveness of the Global Plans; the degree to which objectives have been achieved; and, analyse the link between emergency, rehabilitation and development.

**F** Health, food security, drugs, nutrition, refugees, IDPs, LRRD, coordination, strategy.

**C** Relevance, appropriateness, effectiveness, cost-effectiveness, efficiency, coherence, complementarity, impact.

■ ECHO (October 2001a) *Evaluation of ECHO's Global Plans 2000 and 2001 – Democratic Republic of Congo: Report on the Drug Supply* (Aachen, Germany: ECHO) Germax Gerli GmbH F. von Massow

**O** Assess appropriateness and effectiveness of the Global Plans; the degree to which objectives have been achieved; and, analyse the link between emergency, rehabilitation and development.

**F** Health, drugs, refugees, IDPs, LRRD, coordination, strategy.

**C** Relevance, appropriateness, effectiveness, cost-effectiveness, efficiency, coherence, complementarity, impact.

■ ECHO (October 2001b) *Evaluation of ECHO's Global Plans 2000 and 2001 – Democratic Republic of Congo: Seteur Nutrition/Securité Alimentaire* (Brussels, Belgium: ECHO) M. Logez

**O** Assess appropriateness and effectiveness of the Global Plans; the degree to which objectives have been achieved; and, analyse the link between emergency, rehabilitation and development.

**F** Nutrition, food security, refugees, IDPs, LRRD, coordination, strategy.

**C** Relevance, appropriateness, effectiveness,

cost-effectiveness, efficiency, coherence, complementarity, impact.

■ ECHO (October 2001c) *Evaluation of ECHO's Global Plans 2000 and 2001, Democratic Republic of Congo: Secteur Santé* (Brussels, Belgium: ECHO) O. Barthes

O Assess appropriateness and effectiveness of the Global Plans; the degree to which objectives have been achieved; and, analyse the link between emergency, rehabilitation and development.

F Health, refugees, IDPs, LRRD, coordination, strategy.

C Relevance, appropriateness, effectiveness, cost-effectiveness, efficiency, coherence, complementarity, impact.

■ ECHO (December 2001) *ECHO Global Plan 2000/ Intervention Plan 2001 – Sierra Leone: Synthesis Report* (Brussels, Belgium: ECHO) Simon, P.

O Synthesize the findings of 3 sectoral evaluations of ECHO funded projects in Sierra Leone from September to October 2001.

F Health, nutrition, water, sanitation, child protection.

C Relevance, effectiveness, efficiency, impact, connectedness

■ ECHO (December 2001a) *ECHO Global Plan 2000/ Intervention Plan 2001 – Sierra Leone: Sector 'Uprooted Population'* (Brussels, Belgium: ECHO) Simon, P.

O Evaluate the situation and needs of uprooted populations in Sierra Leone, and projects directly targeting those beneficiaries.

F IDPs, operational issues, resettlement, coordination, LRRD

C Relevance, effectiveness, efficiency, impact, sustainability.

■ ECHO (December 2001b) *ECHO Global Plan 2000/ Intervention Plan 2001 – Sierra Leone: Sector 'Child Protection and War Victims'* (Brussels, Belgium: ECHO) Dal, F.

O Assess projects financed by ECHO for

Child Protection and War Victims in Sierra Leone, and define needs for 2002.

F Child protection, war victims, psychosocial support, reintegration, sexual violence.

C Relevance, effectiveness, efficiency, impact

■ ECHO (December 2001c) *ECHO Global Plan 2000/ Intervention Plan 2001 – Sierra Leone: Health, Nutrition, Water and Sanitation – 2001* (Brussels, Belgium: ECHO) Bousquet, C.

O Assess the suitability of the 2000–2001 ECHO's operations for the health, nutrition, water and sanitation sectors and to make recommendations on the basis of the lessons learned.

F Health, nutrition, water, sanitation.

C Relevance, effectiveness, efficiency, impact, connectedness.

■ ECHO (December 2001d) *Evaluation des Plans Globaux ECHO 2000/2001 – Burundi: Rapport de Synthèse* (Brussels, Belgium: ECHO) Leblanc, M., Barthes, O. and X. Detienne

O Synthesize the results of ECHO's humanitarian activities across all major sectors, between 2000 and 2001, to allow ECHO to decide strategies and improve future actions in Burundi.

F Health, nutrition, food security, water, sanitation.

C Relevance, effectiveness, efficiency, impact, sustainability.

■ ECHO (December 2001e) *Evaluation des Plans Globaux ECHO 2000/2001 – Burundi: Secteur Eau & Assainissement* (Brussels, Belgium: ECHO) Detienne, X.

O Evaluate the results of ECHO's humanitarian activities in the water and sanitation sectors, between 2000 and 2001, to allow ECHO to decide strategies and improve future actions in Burundi.

F Water, sanitation, IDPs, hygiene.

C Relevance, efficiency, effectiveness, impact, sustainability.

■ ECHO (December 2001f) *Evaluation des Plans Globaux ECHO 2000/2001 – Burundi: Sécurité Alimentaire et Production Agricole* (Brussels, Belgium: ECHO) Leblanc, M.

O Evaluate the results of ECHO's humanitarian activities in the food security agriculture sectors, between 2000 and 2001, to allow ECHO to decide strategies and improve future actions in Burundi.

F Food security, agriculture.

C Relevance, effectiveness, efficiency, impact, sustainability.

■ ECHO (December 2001g) *Evaluation des Plans Globaux ECHO 2000/2001 – Burundi: Santé – Nutrition* (Brussels, Belgium: ECHO) Barthès, O.

O Evaluate the results of ECHO's humanitarian activities in the health and nutrition sectors, between 2000 and 2001, to allow ECHO to decide strategies and improve future actions in Burundi.

F Health, nutrition.

C Relevance, effectiveness, efficiency, impact, sustainability.

■ ECHO (December 2001h) *Evaluation of UNHCR Activities funded by ECHO in Serbia, Kosovo, Zambia and Guinea: Synthesis Report* (Brussels, Belgium: ECHO) Prolog Consult Van Bruaene and M. Atkinson

O Assess the set up and the impact of the UNHCR programmes in four country case studies and to see the part played in them by the ECHO contribution.

F Refugees, IDPs, strategic partnership, interagency coordination, protection.

C Relevance, impact, effectiveness, efficiency, sustainability.

■ ECHO (February 2002) *Evaluation of ECHO's 2001 Intervention Plan in Colombia: Assessment of ECHO's Future Strategy in Colombia* (Brussels, Belgium: ECHO) Franklin Advisory Services de Haulleville, A. and Zavala, C.

O Evaluate ECHO's 2001 Intervention Plan and assess its future strategy in Colombia.

F IDPs, operational issues, policy.

C Relevance, effectiveness, efficiency, impact, connectedness

## Food and Agriculture Organisation (FAO)

■ FAO (September 2002) *Thematic Evaluation of Strategy A.3: Preparedness for, and Effective and Sustainable Response to, Food and Agricultural Emergencies* (Rome, Italy: FAO Evaluation Service)

O Assess FAO objectives contributing to strategic objective A3 during the period 1996–2001.

F Preparedness, agricultural relief, LRRD, disaster recovery.

C Relevance, efficiency, appropriateness, effectiveness, impact

## Handicap International (HI)

■ HI (July 2002) *Internal Evaluation of the Post-Drought Emergency Project* (Lyon, France: HI) Villeval, P.

O Assess the suitability of the 2001 drought response in Cambodia and the level to which it has been carried out, whether objectives have been achieved and the impact in term of outputs.

F FFW, food aid, agriculture, technical training.

C Effectiveness, coherence, cost-effectiveness, efficiency, impact.

## International Federation of Red Cross and

### Red Crescent Societies (IFRC)

■ IFRC (May 2002) *Seasonal Flooding: Synthesis of Recent Evaluations and Reviews* (Geneval, Switzerland: IFRC) Mitchell, J.

O Summarize main lessons and findings from recent evaluations, reviews and lesson learning initiatives, that have taken place after recent flood relief operations.

F Preparedness, recovery, rehabilitation, coordination, relief.

C Effectiveness.

■ IFRC (September 2002) *Evaluation of the Goma Volcano Operation* (Geneval, Switzerland: IFRC) Vine Management Consulting Lawry-White, S.

O Assess the Federation's disaster response and

coordination of departments, delegates and Federation members.

**F** Disaster response, management, programming, preparedness, coordination  
**C** Efficiency, effectiveness.

**A** Draws on Red Cross interventions during 2000 and 2001 in Mozambique, Vietnam, Laos, Cambodia, Sudan, India, Bangladesh, Kazakhstan, Venezuela.

### Medecins sans Frontieres (MSF)

■ MSF (November 2001) *Review of MSF El Salvador Earthquake programmes 13/1/01–31/05/01* (Amsterdam, The Netherlands: MSF Holland) Prescott, G.

**O** Evaluate the appropriateness of the MSF earthquake responses and policy following the earthquakes in El Salvador in January and February 2001.

**F** Policy, strategy, coordination, advocacy, public health, management.

**C** Appropriateness, efficiency, impact, cost-effectiveness, coverage, coherence, connectedness.

### Netherlands Ministry of Foreign Affairs

■ Netherlands Ministry of Foreign Affairs (January 2002) *Assistance to Returnees and Vulnerable Groups – Croatia FSN/FSS: Evaluation report* (The Hague, The Netherlands: Netherlands Red Cross) Pupacic, S. and H-J. Bustraan

**O** Evaluate the performance of the 'Community Support Programme' from mid-2000 to mid-2001; assessing impact and effectiveness, and whether the beneficiary criteria were adequate and adhered to.

**F** Beneficiaries, reintegration, rehabilitation, process, refugees, IDPs, management.

**C** Impact, effectiveness, relevance, cost-effectiveness.

### Norway Ministry of Foreign Affairs

■ Norway Ministry of Foreign Affairs (November 2001) *Evaluation of the International Humanitarian Assistance of the Norwegian Red Cross* (Oslo, Norway: MFA Norway)

Channel Research Ltd Brusset, E., Agger, I., Lowery, K-J. and P. Wiles

**O** Describe and assess the performance of the Norwegian Red Cross from 1996–2000 and their role as a channel for support to humanitarian operations.

**F** Coordination, logistics, policy, management, reporting.

**C** Timeliness, efficiency, relevance, effectiveness.

**A** Based on field visits to projects in former Yugoslavia, Rwanda and India.

### Oxfam

■ Oxfam (September 2001) *Internal Evaluation of Emergency Floods Relief and Rehabilitation Programmes 2000–2001* (Dhaka, Bangladesh: Oxfam) Corsellis, T., Rahman, S., Nasreen, M. and H. Ali

**O** Review and assess the effectiveness and impact of the emergency relief and rehabilitation programme in Bangladesh between September 2000 and January 2001.

**F** Floods, shelter, food aid, health, sanitation, coordination.

**C** Effectiveness, impact, appropriateness, coverage, coherence, connectedness, efficiency, cost-effectiveness.

■ Oxfam (March 2002) *Evaluation Report: Oxfam–GB Humanitarian Programme in Bujumbura Rurale* (BUR 191, 199) (London, UK: Oxfam) Smith, R.

**O** Assess Oxfam's work in Bujumbura Rural between 2000 and 2002 and identify key lessons for Oxfam's future work in Burundi.

**F** Water, sanitation, strategy, management, coordination, IDPs.

**C** Impact, effectiveness, appropriateness, cost-effectiveness, sustainability.

■ Oxfam (May 2002) *External Evaluation of Oxfam Brazil's Pernambuco Flood Relief Project: August 2000–May 2002* (London, UK: Oxfam) Roome, R. and O. Rocha

**O** Evaluate the flood response in Brazil between August 2000 and May 2002 to draw lessons for future Oxfam preparation and



operations in flood responses.

**F** Floods, preparedness, housing, health, hygiene, advocacy, policy, LRRD.

**C** Timeliness, appropriateness, effectiveness, coverage.

■ Oxfam (June 2002) *Oxfam GB Angola's Public Health Programme: An evaluation* (London, UK: Oxfam) Sandison, P., Bastable, A. and R. Jantti

**O** Evaluate the programme to formulate lessons learned and recommendations for Oxfam's future public health programme in Angola, based on performance up until April 2002.

**F** Public health, watsan, gender, human resources, management, operational issues.

**C** Impact, effectiveness, efficiency, sustainability, appropriateness.

■ Oxfam (July 2002) *Evaluation Report of Health Component of SLE 110 in Sierra Leone* (London, UK: Oxfam) Lloyd, A.

**O** An evaluation for internal learning and to identify lessons for future programme design and implementation, based on performance from January 2001 to mid-2002.

**F** Health, public health, camps, beneficiary participation, resettlement.

**C** Impact, appropriateness, efficiency, effectiveness, sustainability.

## Tearfund

■ Tearfund (April 2002) *Evaluation Report: South Serbia Shelter Programme, Nis and Presevo Valley Programme, Serbia* (London, UK: Tearfund) Chisholm, M. and J. Brownson

**O** Assess the effectiveness of ECHO-funded 'Shelter Assistance' Project and whether objectives were achieved, between July 2001 and April 2002.

**F** Refugees, IDPs, shelter, LRRD, donor relations.

**C** Appropriateness, effectiveness.

■ Tearfund (May 2002) *Evaluation of XXX in Northern Afghanistan: Final Report* (London, UK: Tearfund) Piper, C.

**O** Evaluate XXX's relief intervention in Northern Afghanistan during May 2002.

**F** Food aid, relief supplies.

**C** Impact, effectiveness.

■ Tearfund (July 2002) *Evaluation of ZZZ Tajikistan and Northern Afghanistan* (London, UK: Tearfund) Anderson, D., Jaeger, T., Kemp, S. and R. Schofield

**O** Assess and give recommendations for improving, the direction and effectiveness of ZZZ's work in Northern Afghanistan and Tajikistan, during July 2002

**F** Micro-credit, shelter, reconstruction, food for work, Food for Asset Creation, health education, sanitation.

**C** Effectiveness.

■ Tearfund (September 2002) *Integrated Food Security Programme in the Communes of Bugabira and Ntega in Kirundo Province: Evaluation Report on the First Phase* (London, UK: Tearfund) Nimubona, D.

**O** Assess the programme's first phase, with a view to improving the performance of the project's second phase.

**F** Food security, agriculture, seed provision.

**C** Relevance, impact.

## United Nations Children's Fund (UNICEF)

■ UNICEF (December 2001) *Evaluation Report: North Sudan* (Khartoum, Sudan: UNICEF) Nginya, P.

**O** Assess the ECHO-funded UNICEF drought emergency response from February 2001 to mid-November 2001.

**F** Water, sanitation, training.

**C** Timeliness.

## United Nations High Commission for

### Refugees (UNHCR)

■ UNHCR (May 2002) *Meeting the Rights and Protection Needs of Refugee Children: An Independent Evaluation of the Impact of UNHCR's Activities* (Geneva, Switzerland: Valid International) Verhey, B. and A. Hallam

**O** Evaluate whether UNHCR is effectively meeting the protection needs of refugee children and the impact of their activities .  
**F** Protection, refugees, children, policy, gender, management tools, operational issues.  
**C** Effectiveness, impact, relevance, coverage.  
**A** Contains an annex identifying lessons-learned from the evaluation methodology and process used by UNHCR and the consultancy team.

### United Nations Mine Action Service (UNMAS)

■ UNMAS (February 2002) *Willing to Listen: An Evaluation of the United Nations Mine Action Programme in Kosovo 1999–2001* (Geneva, Switzerland: UNMAS) The Praxis Group, Ltd Salomans, D., Bullpitt, I., Kannangara, A. and M. Hodgson.

**O** Evaluate the UN-supported mine action programme in Kosovo between early 1999 and the end of 2001.  
**F** Mine action, coordination, management.  
**C** Timeliness, effectiveness, efficiency, cost-effectiveness, impact.

### United Nations Office for the Coordination of Humanitarian Affairs (OCHA)

■ OCHA (April 2002) *An External Review of the CAP* (New York, USA: OCHA) Porter, T.

**O** Provide a status report on the achievements and shortfalls of the CAP as an interagency coordination mechanism, strategic and monitoring tool and as a resource mobilization tool.  
**F** Process of preparing and responding to the CAP.  
**C** Appropriateness, coverage, coherence.  
 Comments: Review is limited to appeals issued in 2000.

### US Agency for International Development (USAID)

■ USAID (April 2001) *Liberia Complex Emergency: Overview of OFDA Programs (1989–1999)* ( Washington

DC, USA: USAID/OFDA) Taddesse, S.

**O** Determine the effectiveness and impact of the OFDA programs and draw lessons for other OFDA programs delivered in a difficult security environment with isolated and inaccessible IDPs.  
**F** Health, nutrition, agriculture, IDPs, strategy, security.  
**C** Effectiveness, impact.

### World Food Programme (WFP)

■ WFP (December 2001) *Full Report of the Evaluation of Ethiopia PRRO 6180: "Targeted Food Assistance for Refugees in Ethiopia and for Refugee Repatriation"* (Rome, Italy: WFP) Kirkby, J., Weingarten, L., Crawford, N. and S. Green

**O** Assess WFP assistance to refugees in Ethiopia between April 2000 and May 2001 and assess the usefulness of the PRRO category.  
**F** Food aid, food security, nutrition, refugees, repatriation, policy, strategy, gender, logistics.  
**C** Relevance, appropriateness, timeliness, efficiency, effectiveness, connectedness.

■ WFP (December 2001a) *Full Report of the Evaluation of Uganda PRRO 6176: "Targeted Food Assistance for Refugees, IDPs and Vulnerable Groups"* (Rome, Italy: WFP) Broughton, B., Jamal, A., Lindholm, J., Maina, W. and D. Tymo

**O** Assess WFP assistance to Uganda between April 2000 and May 2001 and assess the usefulness of the PRRO category.  
**F** Food aid, food security, nutrition, refugees, IDPs, policy, strategy, gender.  
**C** Relevance, appropriateness, timeliness, efficiency, effectiveness, connectedness.

■ WFP (January 2002) *Full Report of the Evaluation of PRRO Somalia 6073.00: "Food aid for Relief and Recovery in Somalia"* (Rome, Italy: WFP) Holt, J. and I. Metaxa

**O** Assess WFP's assistance under the PRRO between January 1999 and July 2001 and the added value of including the operation in the PRRO programme category.

**F** Food aid, food security, coordination, logistics, financing, gender, policy, strategy.

**C** Relevance, timeliness, efficiency, effectiveness, appropriateness, sustainability.

■ WFP (April 2002) *Full Report of the Evaluation of WFP Angola Portfolio* (Rome, Italy: WFP) Bugnion, C., Brusset, E., Pinto, C., Vally, A. and S. Frueh

**O** Assess WFP's portfolio of assistance to Angola between October 1999 and September 2001 and assess the usefulness of the PRRO category.

**F** Food aid, food security, gender, IDPs, policy, logistics, strategy.

**C** Relevance, appropriateness, timeliness, efficiency, effectiveness, connectedness.

■ WFP (April 2002a) *Full Report of the Evaluation of Azerbaijan PRRO 6121.00: Relief and recovery assistance for vulnerable groups* (Rome, Italy: WFP) Kirkby, J., Ibragimbekova, R. and P. Mattei

**O** Assess WFP assistance to Azerbaijan between April 2000 and May 2001 and assess the usefulness of the PRRO category.

**F** Food aid, food security, nutrition, IDPs, policy, strategy, gender.

**C** Relevance, coherence, efficiency, effectiveness.

■ WFP (September 2002) *Full Report of the Evaluation of the GREAT LAKES Regional PRRO 6077 and PRRO 6077.1 (WINGS 10062.0): "Food Aid for Relief and Recovery in the Great Lakes Region"* (Rome, Italy: WFP) Green, S., Schlossman, N., Bugnion, C. and J. Kirkby

**O** Assess the usefulness of the regional PRRO, as a resource mechanism and programming instrument in the Great Lakes, and an effective tool for supporting relief and recovery activities in the region, during February and March 2002.

**F** LRRD, financing, planning, Regional PRRO programming.

**C** Effectiveness, appropriateness, impact.

■ WFP (September 2002a) *Full Report of the Thematic Evaluation of the WFP Commitments to Women – 1996–*

*2001* (Rome, Italy: WFP) El-Solh, C., Philpott, S., Franciosi, P., Trentmann, C., Bessuges, P., Badawi, Z., Hanifa, A., Touré, T., Solano, Y., Alizai, S., Tymo, D., Christensen, I. and J. Fleuren

**O** Assess and evaluate WFP's commitments to Women between 1996–2001, including all policies, measures and programme/project implementation during the period.

**F** Policy, Gender equality, targeting, food distribution, FFW.

**C** Relevance, effectiveness, impact, sustainability.

■ WFP (September 2002b) *Full Report of the Evaluation of IRAN PRRO 6126: Food Assistance and Support for Repatriation of Iraqi and Afghan Refugees in Iran* (Rome, Italy: WFP) Broughton, B., Mattei, P. and M. Kimiagar

**O** Assess WFP assistance under the PRRO in Iran (1999–2002) in order to improve the implementation of the current operation, and assist with planning of the next phase.

**F** Food aid, refugees, gender, environment.

**C** Efficiency, relevance, effectiveness, coherence, sustainability.

■ WFP (September 2002c) *Full Report of the Evaluation of WFP's Special Operations* (Rome, Italy: WFP) ITAD Ltd de Hennin, C. and N. Enqvist

**O** Review justification, planning and processing of WFP Special Operations, and assess performance, impact and results.

**F** Transport, logistics, infrastructure support, air operations.

**C** Effectiveness, cost-effectiveness.

**A** Based on numerous ongoing and recent Special Operations in multiple regions.

### World Health Organization (WHO)

■ WHO (December 2002) *Review of WHO Humanitarian Programs in North Caucasus (Russian Federation)* (Geneva, Switzerland: WHO) De Ville de Goyet and M. Phelan

**O** Examine the role and effectiveness of the WHO emergency programmes to support the health sector in North Caucasus, and

particularly WHO's co-ordinating role, between October and November 2002.

**F** Coordination, information management, management, health.

**C** Effectiveness.

### World Vision

■ World Vision (June 2001) *Evaluation Report: Baringo District Drought Relief Assistance Programme, Kenya* (Missisauga, Canada: World Vision Canada) Borrel, A.

**O** Determine the effectiveness of nutrition interventions and the quality and usefulness of information obtained through rapid nutritional surveys and other information management systems during May 2001.

**F** Nutrition, prevention, human resources, coordination.

**C** Effectiveness.

■ World Vision (December 2001) *Evaluation Report: Emergency Nutrition Response Program, Ethiopia*

(Missisauga, Canada: World Vision Canada) Gardner, A.

**O** Determine the effectiveness of supplemental feeding programmes, March to October 2001, and the quality of information gained through management information systems.

**F** Nutrition, malnourishment.

**C** Effectiveness.

### World Vision/United Nations Children's Fund (UNICEF)

■ World Vision/UNICEF (September 2002) *World Vision/UNICEF Western Afghanistan Emergency Nutrition Response: Final Program Evaluation Report* (Missisauga, Canada: World Vision Canada) Tweddale, A.

**O** Evaluate an emergency nutrition programme during the period January to June 2002 and review programme accomplishments and interventions.

**F** Emergency feeding programmes, nutrition

**C** Effectiveness

## Annex 4 References: Other Sources

- Action Aid (2000)** *ActionAid Accountability Learning and Planning System (ALPS)*. London: ActionAid.
- ALNAP (2001)** *ALNAP Annual Review: Humanitarian Action: Learning from Evaluation*. London: Overseas Development Institute.
- ALNAP (2002)** *ALNAP Annual Review: Humanitarian Action: Improving Performance through Improved Learning*. London: ALNAP. Overseas Development Institute
- ALNAP (2003)** *Global Study. Draft Practitioner's Handbook*. London: ALNAP. Overseas Development Institute (Handbook available mid-2003 mimeo)
- ALNAP (2003)** *ALNAP Global Study: Angola Consultation with and participation by beneficiaries and affected populations in the process of planning, managing, monitoring and evaluating humanitarian action*. London: ALNAP. Overseas Development Institute
- ALNAP (2003)** *ALNAP Global Study: Columbia Consultation with and participation by beneficiaries and affected populations in the process of planning, managing, monitoring and evaluating humanitarian action*. London: ALNAP. Overseas Development Institute
- ALNAP (2003)** *ALNAP Global Study: Sri Lanka Consultation with and participation by beneficiaries and affected populations in the process of planning, managing, monitoring and evaluating humanitarian action*. London: ALNAP. Overseas Development Institute
- American Evaluation Association (1994)** *The Program Evaluation Standards*. Thousand Oaks: Sage/Joint Committee on Standards of Education Evaluation.
- Benini, A (2001)** *Navigating Post-Conflict Environments: Humanitarian Information Management: Decision Support for Mine Action*. Washington: Vietnam Veterans of America Foundation. [[http://vvaf.policy.net/humanitarian/immap\\_report\\_1/immap\\_text.pdf](http://vvaf.policy.net/humanitarian/immap_report_1/immap_text.pdf)]
- Borton, J & Macrae, J (1997)** *Evaluation Synthesis of Emergency Aid Evaluation*. Report 613 London: DFID.
- CIDA (2002)** *Canada making a difference in the world: A policy statement on strengthening aid effectiveness*. Hull: CIDA.
- Collinson, S & Buchanan-Smith, M (2002)** 'International Humanitarian Action and the Accountability of Official Donors', *Briefing Paper 6*. London: Overseas Development Institute/HPG.
- Currión, P (2001)** 'Humanitarian Information Centres: Establishing Coherent Approaches to Field-Based Information Management in Emergencies'. [<http://www.reliefweb.int/symposium/InfoCenters.htm>]
- David, R, & Mancini, A (2003)** 'Transforming Development Practice: the journey in the quest to develop planning, monitoring and evaluation systems that facilitate (rather than hinder) development'. Paper presented at Partners in Development Evaluation - Learning and Accountability, DAC/OECD Workshop, Paris 25–26th March, 2003.
- Davidson, S (2002)** *The Accountable Organization*. Geneva: Humanitarian Accountability Project. [[www.hapgeneva.org](http://www.hapgeneva.org)]
- De Waal, A (1997)** *Famine Crimes: Politics and the Disaster Relief Industry in Africa*. Oxford: James Curry.
- DEC (2002)** *Independent Evaluation of Expenditure of DEC Earthquake Appeal Funds, Volume 2, Full Evaluation Report*. London: Disasters Emergency Committee.
- ECHO (1999)** *Evaluation of the humanitarian aid funded by the European Commission between 01.01.91 and 30.06.96*. Brussels: ECHO.
- ECHO (2001)** *Evaluation of ECHO's Global Plan 2000 – Angola. Water and Sanitation* Brussels: ECHO.

- Edwards, M & Hulme, D (eds) (1995)** *Non-Governmental Organisations: Performance and Accountability*. London: Earthscan/Save the Children.
- Eguren, L (2002)** *An Operational Framework for Field Protection* (Draft) London: PBI European Office [<http://igc.org/pbi/beo>]
- EuronAid (1995)** *Code of Conduct – Food Aid and Food Security*. The Hague: EuronAid.
- FAO (2001)** *Gender and Development Plan of Action, 2002–2007*. Rome: FAO.
- Fetterman, D (2002)** 'Empowerment Evaluation: Building Communities of Practice and a Culture of Learning' in *American Journal of Community Psychology*, 30 (1), pp89–102.
- Geneva International Centre for Humanitarian Mining (2003)** 'IMSMA Web Page' [<http://www.gichd.ch/imsma>]
- Gosling L & Edwards, M (1995)** *Toolkits – a Practical Guide to Assessment, Monitoring, Review, and Evaluation*. Development Manual 5. London: Save the Children, UK.
- Groupe URD (2002)** *Quality Project*, Groupe URD: mimeo.
- Grunewald, F (2001)** 'Responding to long term crises' in *Humanitarian Exchange* (18), pp2–3.
- Hallam, A (1998)** *Evaluating Humanitarian Assistance Programmes in Complex Emergencies*, RRN Good Practice Review. London: Overseas Development Institute.
- HAP (2002)** *Final Report, Afghanistan Field Trial*. Geneva: Humanitarian Accountability Project. [[www.hapgeneva.org](http://www.hapgeneva.org)]
- Hilhorst, D (2002)** 'Being Good at Doing Good? Quality and Accountability of Humanitarian NGOs' in *Disasters* Vol 26, No 3, September. London: Overseas Development Institute.
- HPG (2002)** *International Humanitarian Action and the Accountability of Official Donors, HPG Briefing Paper 6*. London: Overseas Development Institute.
- IANWGE (2001)** *Gender Mainstreaming in the Programme Budget in the United Nations System* New York: Inter-agency Network on Women and Gender Equality.
- IANWGE (2003)** *Gender Mainstreaming in the Programme Budget Process in the UN System. Final Report*. New York: Inter-Agency Network on Women and Gender Equality.
- IASC (2000)** *Gender and Humanitarian Assistance Resource Kit*. Geneva: Inter-Agency Standing Committee.
- IASC (2002)** *Growing the Sheltering Tree - Protecting Rights through Humanitarian Action* Geneva: Inter-Agency Standing Committee.
- ICRC (2001)** *Annual Report*. Geneva: International Committee of the Red Cross.
- ICRC (2001)** *Women Facing War. ICRC study on the impact of armed conflict on women* Geneva: International Committee of the Red Cross.
- IFRC (2002)** *World Disaster Report. Focus on Reducing Risk*. Geneva: International Federation of Red Cross and Red Crescent Societies.
- IRC (2001)** *The IRC Causal Pathway Framework. A Guide to Program Design, Monitoring and Evaluation*. New York: International Rescue Committee.
- JEEAR (1996)** *Joint Evaluation of Emergency Assistance to Rwanda. Humanitarian Aid and Effects Study III of the International Response to Conflict and Genocide: Lessons from the Rwanda Experience*. Copenhagen: JEEAR. Steering Committee of the JEEAR
- Johnson, S & Rogaly, (1999)** *Microfinance and Poverty Reduction*. Oxford: Oxfam.
- Kaiser, K (2002)** *Participatory and Beneficiary-based Approaches to the Evaluation of Humanitarian Programmes*. Geneva: UNHCR.
- King, D & Dillely, M (2001)** 'Structured Humanitarian Assistance Reporting (SHARE)'. [<http://www.reliefweb.int/symposium/SHAREarticle.htm>]

- Lipsev, M (2000)** 'Meta-analysis and the Learning Curve in Evaluation Practice' in American Journal of Evaluation Vol 21 (2), pp207–13.
- LSO (2002)** Report on Workshop for Field Officers of JEFAP Agencies Operating in Central Region. Malawi: ALNAP Learning Support Office.
- Macrae, J (2001)** *Aiding Recovery. The Crisis of Aid in Chronic Political Emergencies*. London: Overseas Development Institute/Zed Press.
- Macrae, J (2002)** 'The 'bilateralisation' of humanitarian response: trends in the financial, contractual and managerial environment of official humanitarian aid, a background paper for UNHCR'. London: Overseas Development Institute
- Macrae, J (2002)** *Uncertain Power: the changing role of official donors in humanitarian action*, HPG Report 12. London: Overseas Development Institute
- Mayhew, B (2002)** 'Finding, Developing and Keeping Programme Managers: a sector wide problem' in Humanitarian Exchange (22) pp 26–8.
- Mehrahtu, E (2002)** 'Perceptions and Practices of Monitoring and Evaluation: international NGO experiences in Ethiopia' in Development in Practice, Vol 12, Nos. 3–4.
- Minear, L (2002)** *The Humanitarian Enterprise. Dilemmas and Discoveries*. Bloomfield: Kumarian Press.
- Morabito, S (2002)** 'Evaluator Roles and Strategies for Expanding Evaluation Process Influence' in American Journal of Evaluation, Vol 23 (3), pp321–30.
- Mosse, D (1998)** *Development as Process: Concepts and Methods for Working with Complexity* London & New York: Routledge.
- MSF-Holland (1999)** *Monitoring in MSF Holland*. Amsterdam: MSF.
- OCHA (2002)** *CAP Review Good Practices Highlights*. New York, OCHA.
- OECD (1999)** *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies* Paris: OECD. [<http://www.oecd.org/pdf/M00021000/M00021478.pdf>]
- OECD-DAC (2001)** *Evaluation Feedback for Effective Learning and Accountability, Evaluation and Aid Effectiveness*. No 5. Paris: OECD-Development Assistance Committee Working Party on Aid Evaluation.
- Patton, M (1997)** *Utilization-focused Evaluation: The New Century Text*. Thousand Oaks: Sage.
- Patton, M (2001)** 'Evaluation, Knowledge Management, Best Practices, and High Quality Lessons Learned' in American Journal of Evaluation, 22 (3), pp 329–336.
- Peppiatt, D, Mitchell, J & Holzmann, P (2001)** 'Cash transfers in emergencies: evaluating benefits and assessing risks', *HPN Network Paper 35*. London: Overseas Development Institute
- Porter, T (2002)** *An External Review of the CAP*. New York: OCHA.
- ProVention Consortium (2002)** 'Lessons learned from recovery after natural disasters: a five country review'. Paper presented at a conference on *The Future of Disaster Risk: Building Safer Cities*. Washington DC: ProVention Consortium.
- Roper, L & Pettit, J (2002)** 'Development and the Learning Organisation: an introduction' in *Development in Practice*, Vol 12 (3&4), pp258–71.
- RRN (1994)** *Code of Conduct for the International Red Cross and Red Crescent Movement and NGOs in Disaster Relief*, RRN Network, Paper 7. London: Overseas Development Institute [<http://www.ODI.org.uk>]
- Shelter Project (2002)** First draft report of the transitional settlement sector. Cambridge: shelterproject.org. [<http://www.arct.cam.ac.uk/shelter/download/drafts/reportdraft1.pdf>]

- Simister, N (2000)** 'Laying the Foundations: the role of data collection in the monitoring systems of development NGOs'. Bath: Centre for Development Studies, mimeo [<http://www.bath.ac.uk/cds//ns2.pdf>]
- SPHERE (2002)** *Institutionalising Sphere. A Great Start, Next Steps: Lessons Learned in Sphere Implementation from 2000–2002*. Geneva: Sphere Project.
- UNHCR (2001)** *Evaluation of the Dabaab Firewood Project, Kenya*. Geneva: UNHCR.
- UNHCR (2002)** *Real-time Humanitarian Evaluations. Some frequently asked questions*. Geneva: UNHCR.
- UNHCR (2002)** *UNHCR Policy on Refugee Women and Guidelines for the Protection: An Assessment of Ten Years of Implementation. An independent assessment by the Women's Commission for Refugee Women and Children*. New York: Women's Commission for Refugee Women and Children.
- UNICEF (2002)** *Managing M & E in Your Country Office: Good Practice Checklist in Stable, Unstable, and Crisis Contexts*. New York: UNICEF.
- UNICEF (2003)** *M&E Training Resource*. New York: UNICEF.
- USAID – Office of Foreign Disaster Assistance (2002)** *Guidelines for Proposals & Reporting*. Washington, DC: USAID.
- Van de Putte, B (2001)** *Follow-up to Evaluations of Humanitarian Programmes*. London: ALNAP.
- Watkins, B (2002)** 'Monitoring the Performance and Impact of Emergency Food Aid: Preliminary Findings from Kenya'. Nairobi: African Evaluation Association Conference paper, mimeo.
- WFP (1998)** *Enabling Development*. Rome: WFP/EB.A/99/4.
- WFP (2000)** *From Crisis to Recovery* Rome: WFP. [[http://www.um.dk./danidaevalueringsrapporter/1997\\_rwanda](http://www.um.dk./danidaevalueringsrapporter/1997_rwanda)]