

**ALNAP Annual Review Series 2001**

# **Humanitarian Action: Learning from Evaluation**

**ALNAP**

Active Learning Network for Accountability and Performance  
in Humanitarian Action

## **Editors**

Rachel Houghton with Kate Robertson

## **Contributors**

**Chapter 1** John Borton with Lucy Carver

**Chapter 2** Tony Beck

**Chapter 3** Based on an expanded version Raymond Apthorpe prepared of his symposium paper at ALNAP's request

**Chapter 4** Tony Beck with John Borton

**About Contributors:** Prof. Raymond Apthorpe, Visiting Professor at Australia National University, Canberra; Dr Tony Beck, Evaluation and Poverty Specialist, Vancouver; John Borton and Lucy Carver, ALNAP Secretariat, London.

© Overseas Development Institute, London, 2001

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

**For further copies** of this publication please contact:

### **ALNAP**

c/o Overseas Development Institute  
111 Westminster Bridge Road  
London SE1 7JD  
United Kingdom  
Tel: +44 (0)20 7922 0300  
Fax: +44 (0)20 7922 0399  
Email: [aln@odi.org.uk](mailto:aln@odi.org.uk)  
Website: [www.alnap.org](http://www.alnap.org)

## Foreword

In April 2000 the full members of ALNAP approved the launch of the Annual Review series. They did so in the belief that an annual summary of evaluation findings, coupled with an assessment of the overall quality of evaluation reports, would provide significant added value to their work and efforts to improve the international humanitarian system's performance. This first Annual Review confirms that belief.

It draws together the principal findings of 49 evaluations of humanitarian action published during 1999 and 2000. The findings highlight areas where the performance of the humanitarian system was judged to have been satisfactory and where poor. The fact that findings are drawn from a large sample of reports gives them considerable force, and they deserve a role in the reflections and dialogues that set the agenda for improving the system's performance.

Synthesising the findings of so many separate evaluations inevitably raises questions about the standards used for assessing performance, and the quality of the evaluation reports themselves. The Annual Review, therefore, combines a synthesis of the evaluations' main findings with a meta-evaluation of the quality of reports, showing us where evaluation practice deserves emulation and where it needs strengthening. By applying a quality proforma developed by ALNAP, the Annual Review has established a precedent. It is not the last word on what constitutes quality, and there will be further debate. However, in providing a systematic basis for assessment, it promises to be an important tool for monitoring the progress of efforts to improve the quality and effectiveness of the evaluation of humanitarian action mechanism.

I firmly believe that this annual opportunity for the humanitarian system to reflect on its performance, and, for those involved in the commissioning, undertaking and use of evaluation to reflect on practice and quality, will prove invaluable. This is a positive addition to the system's growing repertoire of instruments for measuring performance, enabling it to feel proud of its successes and admit to the continuing challenge to improve.

Wayne MacDonald  
Chair of ALNAP  
(Head of Planning, Monitoring and Evaluation, ICRC)  
Geneva, April 2001

# CONTENTS

## Evaluative reports database

### Chapter 1 Introduction

- 1.1 Purpose and scope of the ALNAP Annual Review series
- 1.2 A decade of rapid change
- 1.3 Evaluation: a key tool for accountability and learning
- 1.4 Defining ‘evaluation of humanitarian action’
- 1.5 Evaluation of humanitarian action: current issues

### Chapter 2 Summary of Findings and Assessment of Non-Kosovo Evaluation Reports

- 2.1 Introduction
- 2.2 The sample assessed
- 2.3 Synthesis of main findings
- 2.4 Meta-evaluation of non-Kosovo reports

### Chapter 3 Evaluations of Humanitarian Aid in Response to the 1999 Kosovo Crisis: Synthesis and Meta-evaluation

- 3.1 Overview of this chapter
- 3.2 The Kosovo conflict: a set of interlocking crises
- 3.3 Agency response issues
- 3.4 Meta-evaluation: Kosovo evaluations as learning tools

### Chapter 4 Overview of the analysis and next steps

- 4.1 Introduction
- 4.2 Principal findings and issues raised by the reports
- 4.3 The quality of evaluation reports
- 4.4 Implications of the LRRD Debate for Establishing Evaluation Standards
- 4.5 Next steps in evaluation of humanitarian action
- 4.6 Concluding remarks

## Annexes

- Annex 1** Quality assessment proforma
- Annex 2** Acronyms
- Annex 3** Glossary
- Annex 4** References for source evaluations
- Annex 5** References for other source material

## THE EVALUATIVE REPORTS DATABASE (ERD)

Annual Review principal data source

ALNAP's Evaluative Reports Database (ERD) was set up in 1997 to facilitate access to evaluative reports of humanitarian action and improve interagency and collective learning.

It contains more than 288 evaluative reports of humanitarian action classified under eight categories: Evaluations, Reviews, Lesson Learning Studies, Evaluation Syntheses and Meta-Evaluations, Conference/Seminar and Workshop Reports, Evaluation Follow-Up Studies, Audits and Good Practice Reviews.

At the time of publication, 130 of the reports were categorised under 'Evaluations', the key data source for ALNAP's Annual Reviews.

Although an expanding collection, the majority of the reports held are provided by ALNAP full members and relate directly to humanitarian actions they have either funded or implemented. It is hoped that the Annual Reviews will raise awareness of the ERD resulting in an increased availability of evaluative reports from non-ALNAP sources and a more comprehensive collection.

To encourage the sharing of reports that might otherwise be retained within commissioning organisations, reports may on request be classified 'ALNAP full members only' and password protected within the ALNAP website. If you are aware of an evaluative report of humanitarian action not included, please notify the Database Manager <[alnap@odi.org.uk](mailto:alnap@odi.org.uk)>.

Non-classified reports are open to all, with key sections (e.g., contents; executive summary; principal findings and recommendations; methodology; terms of reference) fully searchable on the ALNAP website <[www.alnap.org](http://www.alnap.org)>.

An additional source is the 'Useful Resources' database. It contains references on key ALNAP areas of interest, such as learning theory and practice, accountability theory and practice, evaluation guidance, evaluation training materials, evaluation theory and practice and humanitarian codes and standards.

# CHAPTER 1

## INTRODUCTION

### 1.1 Purpose and Scope of the Annual Review Series

This publication is the first in the ALNAP Annual Review series and forms part of ALNAP's broader efforts to support and improve learning and accountability within the international humanitarian system<sup>1</sup>. The Annual Review series is intended to:

- provide the humanitarian system with a means to reflect annually on its performance;
- increase awareness of the principal findings of evaluations of humanitarian action published over the previous year;
- monitor, analyse and report on trends in the evaluation of humanitarian action;
- encourage improvements in the quality of evaluation work undertaken, and develop and publicise assessments of the quality of evaluation of humanitarian action reports.

It is also intended to complement other publications focusing on particular aspects of, and issues related to, humanitarian action and the international humanitarian system, such as the *World Disasters Report* (IFRC 2000), *Global Humanitarian Assistance 2000* (IASC 2000) and *The State of the World's Refugees* (UNHCR 2000).

The series will make extensive use of evaluations of humanitarian action (defined later in this chapter) in its analysis, and prove of considerable relevance to a broad spectrum of those directly or indirectly engaged with or within the humanitarian system.

At its best, when not undermined by poor practice and inappropriate resourcing, evaluation of humanitarian action provides an essential tool for the system. It enables organisations and individuals to draw learning from experience through a systematic and impartial scrutiny of performance.

The potential of evaluation of humanitarian action to facilitate system-wide scrutiny and learning becomes all the more powerful when considered and analysed *en masse*. This first Annual Review synthesises 49 such evaluation reports and 5 existing syntheses, published between 1999 and 2000 and held on the ALNAP [Evaluative Reports Database](#) (see p6) – 17 of these focus on humanitarian action in Kosovo.

The story it tells is pertinent to humanitarian policy, advocacy, knowledge management, evaluation and operations. It will equally be of interest to parliamentary and congressional committees that scrutinise public expenditure on humanitarian action; journalists and researchers that focus on humanitarian issues; and those providing training to humanitarian personnel.

In addition, assessment of the quality of the evaluation reports reviewed will be of particular interest to evaluation managers, the independent consultants who undertake evaluations of humanitarian action, and those involved in training evaluation personnel.

### 1.1.1 **Chapter 1: Setting the scene**

This chapter establishes the context for the Annual Review by considering the profound changes that have affected and influenced the international humanitarian system over the past decade, and the subsequent demands for improved mechanisms for accountability and learning. It introduces and defines the ‘evaluation of humanitarian action’, traces the rapid growth in its use, and considers the differences (as well as the similarities) between it and the evaluation of development co-operation.

The chapter reflects on the degree to which evaluation has become ingrained within the international humanitarian system; on the balance between evaluation’s accountability and learning objectives; and on the current gaps in evaluation guidance material. Finally, it introduces ALNAP’s development and application of an evaluation reports quality assessment proforma.

The remainder of the Annual Review is dedicated, through the synthesis and meta-evaluation of the 1999–2000 evaluations, to drawing out and highlighting strengths and weaknesses in current practice, successes and failures, and the dilemmas, lessons and challenges (old and new) facing the system and the evaluation of humanitarian action genre.

To exploit fully the opportunity for analysis provided by the Kosovo set, a series of evaluations resulting from and focussing on humanitarian action within the same context, the synthesis and analysis of the evaluation reports has been split over two chapters.

### 1.1.2 **Chapter 2: the non-Kosovo set**

Chapter 2 reviews the 37 non-Kosovo evaluations. Amongst these are major evaluation efforts: evaluations of ECHO’s operations 1991–1996 and 1996–2000, and Danida’s humanitarian assistance 1992–98; evaluations of responses to major humanitarian emergencies such as Central American countries affected by Hurricane Mitch in 1998; flooding in China and Bangladesh in 1998–99; food insecurity in North Korea during the late 1990s; civil conflict in East Timor in 1999; and evaluations of humanitarian action in continuing conflict countries such as Afghanistan, the Democratic Republic of the Congo, Sudan and the Caucasus.

### 1.1.3 **Chapter 3: the Kosovo set**

The international response to the humanitarian needs created by the 1999 Kosovo conflict ranks as one of the largest, in terms of the scale of resources. It is also one of the most contentious, with many nations involved in the provision of humanitarian assistance also involved, via their military forces and NATO, as protagonists in the conflict between NATO and the Federal Republic of Yugoslavia. This has been the subject of more than 20 separate evaluations of humanitarian action of which 16 have been published or made available to ALNAP.

Chapter 3 is dedicated to the ‘Kosovo’ evaluation reports, enriching its analysis through reference to some 50 other Kosovo evaluative reports. It presents a synthesis and

comments on the quality of the evaluation reports in terms of their contribution to institutional learning. The analysis is also informed by discussions at the ALNAP *Learning from Evaluation: Humanitarian Assistance and Protection in Kosovo* symposium, in Geneva in October 2000<sup>2</sup>.

#### 1.1.4 Chapter 4: Conclusions

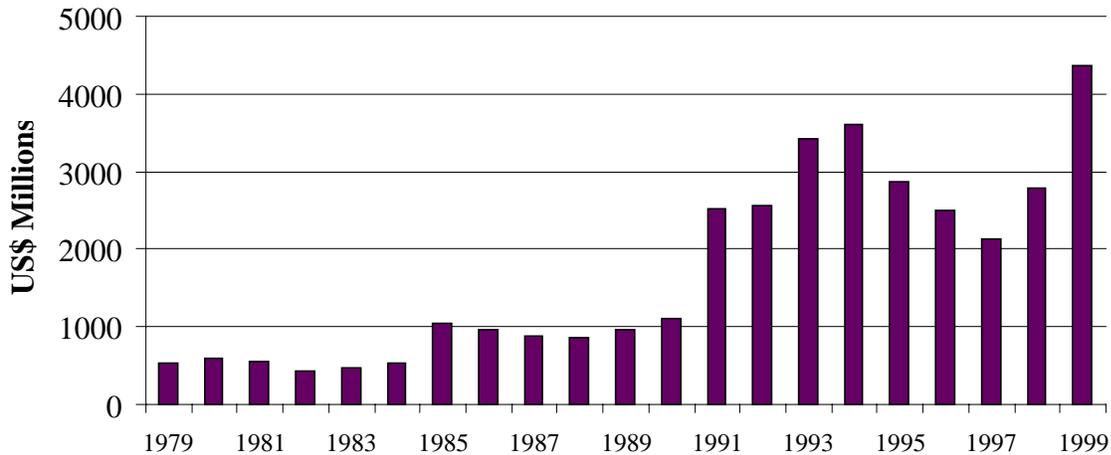
Chapter 4 draws together and reflects on the findings of the preceding chapters. In so doing, it outlines an agenda of issues to be addressed by the international humanitarian system and by those involved in the commissioning, undertaking and use of evaluations of humanitarian action. Based as it is on 49 evaluations published over a two-year period covering operations in 36 countries in 4 regions of the world, the agenda it sets out is robust and comprehensive.

## 1.2 A Decade of Rapid Change

The publication of this first ALNAP Annual Review coincides with the tenth anniversary of Operation Provide Comfort<sup>3</sup>, an unprecedented military intervention and humanitarian operation launched in April 1991 to provide assistance to displaced Kurds trapped in the mountains along the Iraqi/Turkish border. The operation contributed to a sharp increase in the overall annual expenditure on humanitarian assistance (see [Figure 1.1](#)) and marked the start of a decade of frenetic activity for the international humanitarian system. High levels of expenditure were sustained throughout and, boosted by the Kosovo operations, in 1999 ended at its highest level ever.

This ten-year period witnessed an unprecedented number of large-scale humanitarian operations (notably Somalia, southern Africa, former Yugoslavia, Rwanda, Central America, Kosovo, and East Timor) many involving armed intervention in the form of peace enforcement or peacekeeping operations. Humanitarian assistance shifted from a focus primarily on refugees in asylum countries to a post-Cold War regime characterised by the provision of assistance to populations within areas of ongoing conflict.

**Fig. 1.1 Humanitarian Assistance in Real Terms (1998 prices and exchange rates)**



Source: IASC, 2000 (updated from OECD DAC figures 2001)

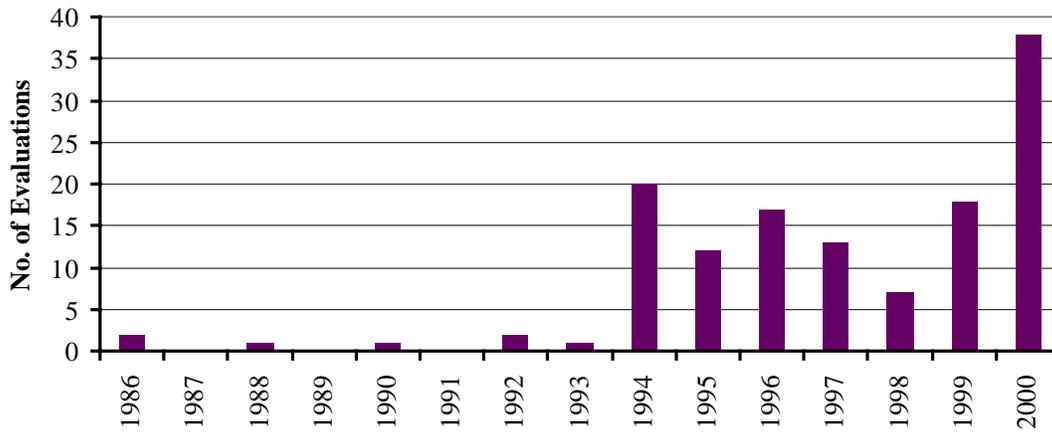
It also saw an unprecedented debate on humanitarian policies and practices and the development of fundamental critiques of humanitarianism (e.g. de Waal, 1997), debates and critiques that continue today. How should humanitarian assistance be provided in contexts where it is often impossible to differentiate civilians from soldiers and where assistance can contribute to the fighting ability of parties to the conflict? What is the appropriate relationship between humanitarian organisations and military contingents engaged in peacekeeping or peace-enforcement? How can the safety of humanitarian staff working in areas of active conflict be ensured?

It is important to recognise that the developments within the international humanitarian system took place against a backdrop of fundamental change in the management of national and international public sector organisations, resulting from the ideology of neo-liberalism. Behind the political rhetoric of downsizing government, increasing quality and value for money in public services, programmes were introduced under headings such as 'total quality management' or 'results-based management'. These aimed at improving the clarity of objectives and the measurement of processes and results (see Kirkpatrick and Martinez Lucio, 1995; Wholey, 1997). Though contentious, such programmes did contribute to an increased transparency of processes and, to an extent, of organisational accountability in the public sector.

Similar programmes have permeated through to the international humanitarian system as it has also attempted to take on board the new management approaches. A significant source of change has been the bilateral donor organisations and the requirements attached to their

funding contributions to multilateral organisations, UN agencies and non-governmental organisations (NGOs).

**Fig. 1.2 Evaluations in ERD by year of publication**



Responding to these new demands and pressures, the system has in the past decade instigated and witnessed a range of initiatives aimed at improving its own transparency and accountability. These include the development of codes of conduct, benchmarks and standards and the exploration of mechanisms for encouraging compliance with these codes and standards (see Box 1.1). Significantly, it was a system-wide evaluation of the international humanitarian response to the genocide in Rwanda (JEEAR, 1996) that provided a final impetus for some of these, including ALNAP.

### **Box 1.1 Principal Accountability Initiatives in the Humanitarian System over the Last Decade**

**1994** Publication of the code of conduct for the International Red Cross and Red Crescent movement and NGOs in disaster relief. [www.ifrc.org](http://www.ifrc.org)

**1996** Publication of the Joint Evaluation of Emergency Assistance to Rwanda. [www.um.dk](http://www.um.dk)

Start of the Sphere Project by a coalition of European and US NGOs to develop minimum standards for five sectors of assistance which resulted in the publication of a final version of a *Humanitarian Charter and Minimum Standards in Disaster Response* in 2000. [www.sphereproject.org](http://www.sphereproject.org)

Formation of People in Aid by a group of UK organisations to focus on issues in the management and support of aid personnel and which in 1997 produced the People in Aid Code of Best Practice. [www.peopleinaid.org.uk](http://www.peopleinaid.org.uk)

**1997** Formation of the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) to provide an interagency forum to discuss and link these initiatives and undertake complementary activities on learning and accountability issues. [www.alnap.org](http://www.alnap.org)

Group of UK organisations forms to develop the concept of a Humanitarian Ombudsman (renamed in 2000 as the Humanitarian Accountability Project) as recommended by the Rwanda evaluation – in 2001 this initiative was raised to an international action-research plane. [www.oneworld.org/ombudsman](http://www.oneworld.org/ombudsman)

### **1.3 Evaluation: a Key Tool for Accountability and Learning**

The changes and pressures of the 1990s, combined with wider management changes, resulted in an unprecedented level of scrutiny of the system, both by outsiders, and from within. Bilateral donor organisations began to apply their new management approaches to their funding and therefore to those they funded, but these management strategies were

complemented by growing recognition across the system of the need to learn from experience and translate learning into improved practice.

A powerful spur to this new-found commitment to improving learning and performance was the increased competition for funds, coupled with the realisation that greater accountability and transparency were necessary to maintain the support of bilateral and multilateral donors.

A wide range of mechanisms have been involved in both scrutiny and learning, including commissions of enquiry, parliamentary committees, evaluations, reviews, learning workshops, media coverage and research studies. While all play a useful role for the international humanitarian system, evaluation has emerged as a key tool for assessing how effectively resources have been used and what lessons might be learnt.

### **1.3.1 Evaluation of humanitarian action**

Evaluation has been variously defined, but its central concepts are those described by Scriven in his *Evaluation Thesaurus*: ‘the process of determining the merit, worth or value of something, or the product of that process,’ (Scriven, 1991: p139). The application of evaluation to particular areas of activity has resulted in more specific definitions, reflecting the particularities of those activities. The definition of evaluation of humanitarian action is introduced in [Section 1.4](#).

### Box 1.2 The Development of Evaluation

As a field of professional practice, evaluation began in the United States of America during the 1960s when the ‘Great Society’ legislation poured Federal expenditures into poverty alleviation, housing, welfare and education programmes. Many of these programmes were required by legislation to undertake evaluations of their effectiveness to guide use of available funding. Evaluation then spread beyond the Federal programmes and was applied to other areas of US public expenditure and then adopted as a practice in other countries. As the use of evaluation grew and spread, so professional evaluation societies were formed, with the first International Evaluation Conference being held in 1995.

Evaluation was applied to international development assistance at a comparatively early stage. The decolonisation process of the 1950s and 60s led to new departments and ministries for the administration of development cooperation programmes. Evaluation units were gradually introduced and by the end of the 1970s most aid-administering organisations had one in some form or other. Under the auspices of the Organisation for Economic Cooperation and Development (OECD), representatives of these units formed a Group on Aid Effectiveness that subsequently evolved into the DAC Working Party on Aid Evaluation. It meets regularly to share experience, to improve evaluation practice and to strengthen its use as an instrument of development cooperation policy.

[www.oecd.org/dac/evaluation](http://www.oecd.org/dac/evaluation)

Compared with the application of evaluation to development cooperation (see [Box 1.2](#) above), its application to humanitarian action has been slower. This was partly the result of initial resistance but also of technical, practical and methodological difficulties (discussed later). The first evaluations of humanitarian action weren’t undertaken until the second half of the 1980s. On the evidence available, 1993 appears to have been the year in which the evaluation of humanitarian action took off (see [Figure 1.2](#)). Since then, however, a boom has been underway in the use of evaluation by the humanitarian system. Interestingly the sharp increase in the number of evaluations in 1993 appears to follow on from the 1991 funding increase (see [Figure 1.1](#)) with an approximately two-year time lag.

What does this boom represent and how long can it be expected to last? Is it bringing genuine improvements to the accountability and learning processes within the international humanitarian system? These questions are central to this Annual Review.

The boom undoubtedly represents a significant investment by the humanitarian system, and presents a considerable opportunity for critical reflection and learning in humanitarian operations. While information on the overall level of resources devoted to evaluation of

humanitarian action is not readily available, evidence from a benchmarking study, being undertaken by the ICRC<sup>4</sup> at time of publication, is indicating that humanitarian agencies devote on average the equivalent of 0.5% of their humanitarian action expenditure to inspection, audit and evaluation activities.

However, if evaluation is to continue to receive its current levels of attention and resourcing, and be embraced by all – whether at policy or operational level – it needs to improve on the quality of the undertaking and its product to demonstrate clearly its contribution to improved performance.

### 1.3.2 **The ALNAP Evaluative Reports Database (ERD)**

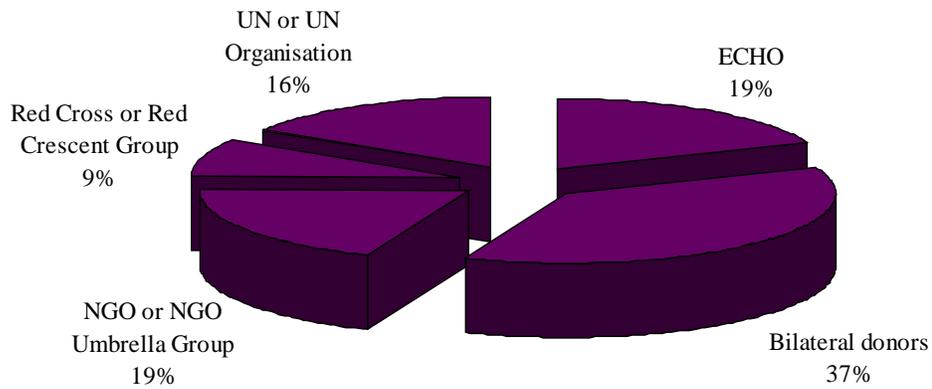
From the outset, those involved in the formation of ALNAP (in 1997) recognised the actual and potential role of evaluation, as well as the need for a mechanism that would enable the systematic collection and sharing of evaluative reports to facilitate and improve interagency and collective learning. The ERD<sup>5</sup> has been a central ALNAP activity ever since, and the Annual Review takes the initiative one step further. Using the ERD as its primary source the Review provides a baseline analysis against which to track and report on future trends. (see [Box 1.3](#)).

For the full accountability and learning potential of evaluation to be realised, a more systematic exploitation of the mechanism will be required.

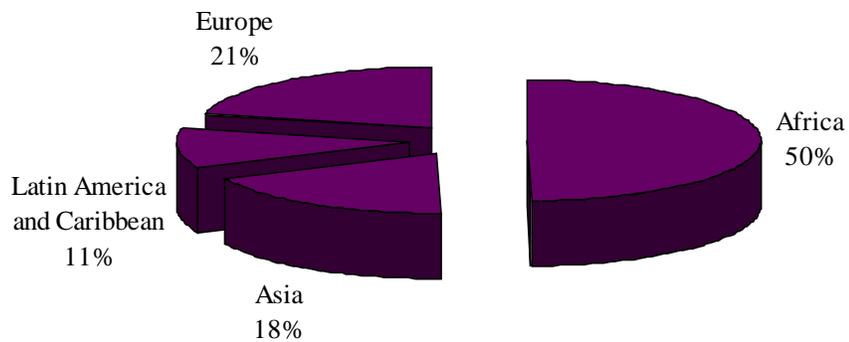
### Box 1.3 The Evaluative Reports Database: Trends

Analysis of the 130 reports categorised as Evaluations of Humanitarian Action in the ERD:

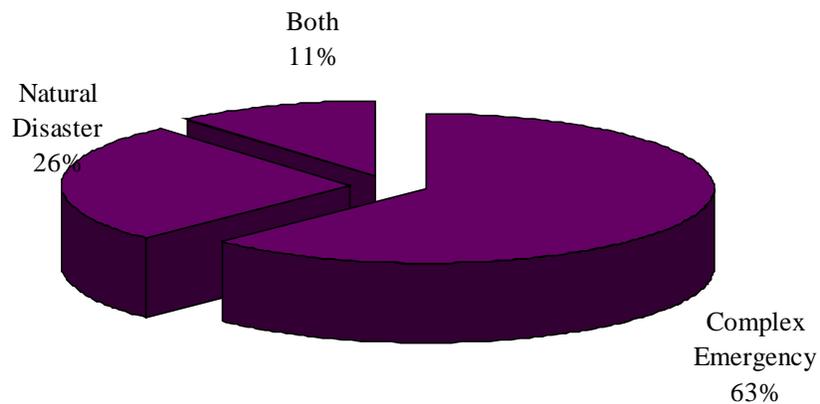
#### Commissioning Organisations



#### Geographical Coverage



#### Type of Emergency



## 1.4 Defining 'Evaluation of Humanitarian Action'

Defining the key characteristics of 'evaluation of humanitarian action' was of particular importance in determining the reports to be considered in the Annual Review. Using DAC's definition as a starting point, ALNAP identified key characteristics, resulting in the adoption of the following in early 2001:

'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 e.g. efficiency, effectiveness/timeliness/coordination, impact, connectedness, relevance/appropriateness, coverage, coherence and, as appropriate, protection (see OECD-DAC, 1999 for an explanation of these criteria).
- It articulates findings, draws conclusions and makes recommendations.'

Other than in Chapter 3, which also draws on other types of evaluative report looking at Kosovo, the Annual Review considers only those reports that conform to the above definition. (The process of categorising the reports is undertaken only on the basis of information contained in, or that can be inferred from, a report.)

While adherence to the definition has the considerable benefit of providing a coherent set of reports, it inevitably excludes. Such exclusion should not be read as a negative judgement on the quality or value of other types of evaluative report but simply as an acknowledgement of their differences. As the approach taken in Chapter 3 indicates, other types of evaluative report can make a valuable contribution to accountability and lesson learning. The process of comparing and assessing the value and contribution of evaluation and other types of evaluative reports also helps to sharpen analysis of the strengths and weaknesses of each.

### 1.4.1 Particularities of evaluation of humanitarian action

The evaluation of humanitarian action should be viewed as a particular and recent form of evaluation despite drawing significantly on the techniques and approaches developed in the 1970s for the evaluation of international development co-operation. Both provide reports intended to feed into the policy process and improve future performance, both are centrally concerned with assessment of impact, appropriateness, effectiveness etc, and both are generally undertaken by external consultants or mixed teams. However, the major contextual differences between the humanitarian and development arenas have required considerable adaptation of the development genre and the adoption of particular criteria, techniques and approaches for the former.

The principal contextual differences are indicated below.

- Humanitarian programmes frequently operate under considerable time pressure and in very fluid and fast-changing contexts. Consequently, project cycles are compressed so that initial planning and baseline data gathering, to enable ‘before’ and ‘after’ comparisons, are so perfunctory (if done at all) as to be of little use in subsequent evaluations. This lack of baseline data, and in some cases basic project information, requires evaluators of humanitarian action to reconstruct ‘histories’ and ‘pictures’.
- Humanitarian programmes often operate in areas where the police, judiciary and press are overwhelmed, ineffective or no longer functioning. This increases both the pressure on, and the significance of, the evaluations’ accountability role.
- Humanitarian programmes, with their focus on saving lives in extreme situations, are often accorded a high profile by the media and politicians. The resulting increase in sensitivity, to direct or implied criticism of programmes under evaluation, impacts both on the process and how results are placed in the public domain.
- Humanitarian operations invariably involve a number of different agencies, in contrast to the ‘dominant and clear leader’ usually found in development contexts. This multiplicity is often characterised by a high degree of interdependency, where one agency’s performance relies significantly on the performance of others – e.g. for funding or delivery in the delivery chain, or the provision of complementary services. This makes trying to attribute relative roles and evaluate shortcomings and impacts more difficult.
- The overlap in the criteria used by the development co-operation and humanitarian genres of evaluation is acknowledged, but some are less applicable than others. For instance, sustainability<sup>6</sup> is often not appropriate for interventions with short-term objectives. Criteria such as coherence and coordination are of central importance in evaluating humanitarian action, given the need for policy coherence in respect of security, diplomatic, humanitarian and development actions.
- Humanitarian intervention is often highly multidisciplinary – e.g., public health, logistics, food and nutrition, water and sanitation, etc. Evaluation teams are required to reflect such diversity in their composition and expertise.
- Conflicts polarise perspectives so that the same events are often subject to widely differing interpretations, diminishing the space for objective assessment.
- The experience of conflict or a natural disaster may traumatise individuals who would have acted as information sources for the evaluation.
- Finally, teams undertaking evaluations of humanitarian action often experience the same problems (insecurity, logistical difficulties, increased risk of illness, etc.) as the agency personnel whose activities they are assessing.

## 1.5 Evaluation of Humanitarian Action: Current Issues

It is useful at this stage to introduce four issues currently facing the system. These have emerged from experiences and discussion within ALNAP and an analysis of the ERD:

- the extent to which evaluation is now ingrained in the humanitarian system;
- the balance between the dual aspirations of evaluation – accountability and lesson learning;
- the gaps in existing guidance on the evaluation of humanitarian action;

- the assessment of quality in relation to evaluation reports.

### 1.5.1 **Is evaluation ingrained in the humanitarian system?**

Statistics drawn from the ERD (see [Box 1.3](#)) need to be interpreted with caution. While bilateral and multilateral donors commissioned 54% of the total number of evaluations held, the focus of these is usually on the performance of NGOs and to a lesser extent that of UN and Red Cross agencies, rather than on their own performance.

Despite the need for caution, evidence suggests that not all parts of the international humanitarian system are contributing to the same degree to the evaluation of humanitarian action 'boom'. Bilateral and multilateral donor organisations appear significantly better at commissioning and sharing their evaluations than other types of organisation. The UN group accounts for just 17% of the total. However, the UN is able to point to the declassification, in 2000, of all UNHCR's evaluation reports as a major achievement.

NGOs were responsible for commissioning 20% of the evaluation reports available on the ERD, however no less than 54% of these reports were commissioned by just two organisations (MSF-Holland and the UK Disasters Emergency Committee [DEC]). The available data does not show conclusively whether these results reflect a reluctance to share evaluation reports, a general lack of commitment to the evaluation process, or a tendency to use internal personnel for evaluative reviews, thus falling outside the ALNAP definition.

The dependence of most NGOs on private funding sources, particularly susceptible to influence by adverse publicity, is probably a significant factor contributing to NGOs' possible caution and sensitivity in the commissioning and/or sharing of evaluations. It is not however in the interests of the international humanitarian system that a group of agencies that plays such a significant role in the delivery of assistance to affected populations does not engage more directly and openly with such an important vehicle for accountability and learning. Greater encouragement may need to be given to NGOs to undertake evaluations of humanitarian action and share the reports not only with other agencies, but more importantly with their supporters and the wider donating public.

While the extent to which the evaluation mechanism has become part of the culture of the international humanitarian system is important in itself, it remains of limited significance if evaluation findings, conclusions and recommendations are not being systematically acted on. The results of work commissioned by ALNAP on the follow-up to evaluation will feature in some detail in the second Annual Review, to be published in April 2002.

### 1.5.2 **Balance between accountability and lesson-learning**

The tension between the principal dual objectives of evaluation of humanitarian action, accountability and lesson-learning, has long been recognised. For the most part, it is regarded as a creative tension that contributes to the uniqueness and potential value of the evaluation process.

The view held by this author is that lessons are more likely to be learned if the process is open to external scrutiny with organisations held accountable for their actions (or

inactions). However, the presence of the accountability objective, enabling ‘outsiders’ to assess the performance of organisations and their staff (who tend to see themselves as doing their best in adversity), is also acknowledged as a major factor in resistance to evaluation of humanitarian action.

The relative balance accorded these two objectives influences the evaluation approach adopted. An evaluation placing emphasis on accountability tends to be perceived as a threat to those being evaluated. This often results in a process with less engagement between the evaluators and personnel from the agency being evaluated than an evaluation emphasising learning. The latter tends to take, and arguably requires, a more participatory approach, engaging both programme staff and its intended beneficiaries. The choice of emphasis, however, remains with the commissioning organisation and, though some will have strong accountability requirements imposed on them by their statutes and governing bodies, others’ accountability requirements will be lower, perhaps by virtue of their funding structures or weak relations with funders. Many NGOs fall into this category, allowing them greater leeway to emphasise the lesson-learning objective in their approach to evaluation.

Given the intensity of humanitarian action over the last decade and the need for the humanitarian system to undertake fundamental changes in response to the sharper critiques, some may be tempted to argue that evaluations of humanitarian action should place greater emphasis on the lesson-learning objective. However, such a general recommendation would not be appropriate, as different organisations and contexts will require different approaches and emphases. Without the encouragement and degree of pressure afforded by the accountability objective, an agency’s commitment to learning the lessons drawn, and integrating them into its policies and practices, may well be insufficient to drive the necessary changes and realise the goal of improved performance.

While there are many genuinely new issues (e.g., protection and security) a striking feature of the analysis in the following chapters is the extent to which the old issues recur. Among them poor interagency coordination (including that between donor organisations); poor linkages between the humanitarian and development system; and, the lack of beneficiary participation in humanitarian action and its evaluation. Perhaps the lesson-learning objective should be given emphasis only in those areas where there are genuinely new lessons to learn whilst accountability should be given emphasis in other areas where lessons are evidently not being learnt, or not being translated into improved practice.

Whatever the chosen emphasis, a fundamental and critical factor in good evaluation practice is that prioritisation of objectives should be shared by key stakeholders and made explicit in the terms of reference and throughout the evaluation process.

The ALNAP book *Evaluating International Humanitarian Action: Perspectives from Practitioners* (Wood et al., 2001) concludes that lack of clarity on the purpose and focus of evaluations of humanitarian action has, together with inadequate time and resourcing, been a significant factor limiting the effectiveness and value of evaluations:

‘Ensuring that the evaluation team knows what is wanted and for what purpose, and forcing the commissioning agency, and not just its evaluation unit, to think about this and clarify the usually excessive and even incoherent TOR is a critical step to achieving more effective and useable evaluations.’

### 1.5.3 Gaps in guidance

Developments in thinking and practice around evaluation of humanitarian action are so rapid that what might be regarded as the ‘first generation’ of guidance (Hallam, 1998; OECD-DAC, 1999) is already considered to be in need of updating.

ALNAP has identified guidance gaps in a range of areas and prioritised the development of practical ‘how to evaluate’ guidance on the following:

- Protection
- Policy
- Security
- Human resources
- Information & knowledge management
- Process (how does one evaluate against recognised criteria such as ‘coherence’, ‘connectedness’ and ‘coverage’?)

Progress will be reported in the second Annual Review to be published in April 2002.

### 1.5.4 Assessing the quality of evaluation reports

To some extent, the credibility, potential and effectiveness of evaluation as a tool for change, as well as the follow-up to evaluation, is being undermined by the variable quality of both the undertaking and its product. A pressing requirement therefore is to reach general agreement on what constitutes good practice in the evaluation of humanitarian action and to develop quality benchmarks for use by evaluation managers and evaluators.

In October 2000 ALNAP full members triggered a process to identify and agree appropriate criteria for assessing the quality of evaluation reports of humanitarian action. A preliminary ‘quality’ proforma<sup>8</sup> (see [Annex 1](#)) was developed drawing on existing guidance, including the OECD-DAC guidance, and a growing body of what is increasingly acknowledged as good practice in evaluation. It uses the following 11 headings:

- Purpose and focus of the evaluation
- Constraints experienced
- TOR, team composition and time allowed
- Information on the context and the intervention
- Methodology and transparency
- Consultation with beneficiaries and the affected population
- Reference to international standards
- Attention to gender and vulnerable or marginalised groups
- Coverage of factors potentially influencing performance
- Conclusions and recommendations
- Legibility and dissemination of the final report

The first two and their sub-headings allow key contextual information to be established, and the nine other heading areas contain a series of criteria against which quality assessments of the evaluation reports considered in this Annual Review have been undertaken. Assessment findings and analysis thereof appear in aggregate form in [Chapter 2](#), in respect of the non-Kosovo evaluation reports, and in [Chapter 4](#) in respect of all the evaluation reports considered in this Annual Review.

Having set the background, we can move on to consider the synthesis and meta-evaluation of the evaluations of humanitarian action reports published during 1999–2000 for the non-Kosovo and Kosovo sets.

# CHAPTER 2

## SUMMARY OF FINDINGS AND ASSESSMENT OF NON-KOSOVO EVALUATION REPORTS

### 2.1 Introduction

This chapter focuses on two interlinked themes:

1. synthesis of the main findings, conclusions and recommendations of non-Kosovo 1999–2000 evaluation reports of humanitarian action;
2. meta-evaluation of non-Kosovo 1999–2000 evaluation reports.

The purpose of the synthesis is to provide an overview of the performance of the humanitarian system for the period under review. The topics covered mirror the main topics in the evaluation reports themselves, they also represent the main issues in humanitarian action as defined by OECD-DAC (1999) and the general evaluation and humanitarian literature. The meta-evaluation assesses the quality of the evaluation reports. The preliminary proforma was developed to assist in this objective. This attempts to measure the quality of reports against generally acknowledged good practice criteria. The proforma was introduced in [Chapter 1](#) and further details are provided in the meta-evaluation section of this chapter.

### 2.2 The Sample Assessed

This chapter analyses the findings of 37 evaluation reports of humanitarian action completed during the 1999–2000 period that did not cover Kosovo. These constitute most of the major evaluation reports for the period. All the reports are classified as ‘evaluations of humanitarian action’ in the ALNAP Evaluative Reports Database. The set comprises 33 individual evaluation reports and four synthesis reports which themselves cover a further 237 evaluations<sup>1</sup>. Consequently this chapter represents one of the most comprehensive surveys of such evaluations to date.

The four synthesis reports are:

- DANIDA (1999a), which covers 7 evaluations (also assessed individually in this chapter);
- ECHO (1999), which covers 140 evaluations;
- ECHO (2000, 2000a), which covers 93 evaluations;
- USAID (2000a), which covers 4 evaluations.

The 33 individual evaluation reports are listed together in Annex 4, where they are broken down by salient characteristics. Of these, 14 relate to complex emergencies, 11 relate mainly to rehabilitation after conflict or complex emergencies, and nine relate to natural disasters. The total disbursement evaluated by the individual evaluation reports amounts to some US\$1.7bn.<sup>2</sup>

About half of the evaluation reports explicitly note that they have a joint lesson-learning and accountability focus; also five have a mixed focus on lesson-learning and planning. (All conform to the ALNAP definition of the evaluation of humanitarian action and are judged to have lesson-learning and accountability objectives, even where not explicitly stated). The majority of the reports evaluate both relief and rehabilitation, and many cover the connection between the two. The set covers 30 countries and 4 regions, including Afghanistan, Central America, the Great Lakes and the Horn of Africa regions, Angola, Bangladesh, China, East Timor, India, North Korea, and the Caucasus. It also covers a wide range of interventions, the most common of which were related to nutrition, health, water and shelter (see [Annex 4](#)). The humanitarian actors assessed are for the most part the main UN agencies (WFP, UNHCR, ICRC, IFRC, IOM, OCHA and UNICEF) and the major INGOs and their in-country partners. Due to the wide variety of implementation channels, individual reports mainly cover both UN agencies and INGOs.

It is difficult to generalise when dealing with such a diverse set of evaluation reports and comparing across a range of emergencies and countries. However, a number of common areas were found and these are highlighted below. Findings are also compared to those of earlier synthesis studies. Unless otherwise noted, findings are consistent between the individual and synthesis reports.

All the evaluation reports held by ALNAP in English for the 1999–2000 period have been included. Representativeness was not an issue in terms of selecting which ones to review. ECHO, as one of the largest agencies, is adequately represented by eight studies. However, one bilateral agency (DANIDA), is comparatively over-represented, and others (JICA, USAID, GTZ) are under-represented in terms of the number of reports considered in relation to the organisations' overall expenditures and roles within the international humanitarian system. The issue of representativeness could therefore become problematic if the findings of the chapter are used as a baseline against which to measure future progress. Representativeness in terms of multilateral agencies and INGOs, both of which are assessed directly in individual reports and as part of bilateral agency sponsored evaluations, is considered to be adequate. Of the 33 individual evaluation reports, all were commissioned by the agency funding the intervention: 10 by bilateral agencies, 9 by INGOs or INGO consortia, and 14 by multilateral agencies (see [Annex 4](#)).

## **2.3 Synthesis of Main Findings**

### **2.3.1 Summary**

The overwhelming majority of evaluation reports conclude that short-term relief efforts are achieving their purpose and meeting the objectives set for the intervention. These objectives are mainly phrased in terms of saving lives at risk, and distributing food, water and basic health care services. Almost all short-term relief is considered to be relevant to needs, and the level of resources devoted to short-term relief is also considered appropriate. Major ongoing weaknesses are highlighted, in particular the failure of coordination and

reporting. Similar findings from other synthesis studies can be found in Apthorpe (1997) and Borton & Macrae (1997). One question that arises from these findings is whether the objectives established by agencies for short-term relief are themselves adequate and appropriate. This is discussed below in the sections on protection and impact.

When interventions are assessed against longer term rehabilitation objectives, the picture is mixed. A general conclusion drawn by this analysis is that the further humanitarian action moves from relief the less likely it is to be successful. Part of the reason for the decrease in success appears to be confusion and disagreement concerning the links between relief, rehabilitation and development.

The overall positive findings noted above in relation to short-term relief, however, need to be read in the context of overall weaknesses apparent in evaluation approaches and methodologies (see, in particular, [\*Methodology and transparency\*](#)). While relatively strong on contextual background, terms of reference, conclusions and recommendations, and legibility, evaluation reports are insufficient with regard to clarity in presentation of methodology and the methodology itself, attention to international standards, and attention to gender. Weakness in these areas calls into question the credibility of results of the evaluations. Current gaps may be filled by more rigorous gate-keeping functions carried out by commissioning agencies, including more comprehensive terms of reference, which need to be enforced. Furthermore, findings might have been less positive if the methodological approach of the evaluations had been more rigorous.

### 2.3.2 **Impact, effectiveness and connectedness<sup>3</sup>**

The findings of the evaluations on impact, effectiveness and connectedness are considered together here as these criteria are closely linked. Impact and connectedness in particular are often considered as key measures of an intervention's performance. As noted above, if evaluated in terms of their own objectives (i.e., effectiveness) the actions were largely successful. The following comments are representative: 'In the area of intervention, diarrhoea outbreaks occurred, but were effectively controlled by adequate response by MSF-H. The case fatality rate remained below the target set,' (MSF, 1999: p13); and, 'DEC agencies, despite severe practical difficulties, were effective in responding to the crisis in areas controlled by the Government of Sudan,' (DEC, 1999: section 8.6). Food aid and the provision of health and water services were found to be mainly successful. As the DANIDA synthesis study points out: 'At an operational level, there have been significant improvements in the quality of humanitarian response over the last decade. ... The quality of [DANIDA] intervention was generally high,' (DANIDA, 1999a: p viii).

### Box 2.1      **Successful Impact?**

‘Despite incomplete geographical coverage due to the limited capacity of the agencies, the emergency water and sanitation needs in the main population centres have been addressed. There have been no outbreaks of disease and no serious water shortages reported. Credit for the rapid provision of emergency healthcare across the country (to district level at least) such as restarting clinics, referral services and undertaking the screening of returnees, can be attributed to the quick deployment of ICRC and the INGOs. UNICEF’s timely vaccination campaign also contributed to the avoidance of major epidemics. Early food distributions averted a food emergency. Since then, the distribution of food supported nearly all other sectors in reaching their respective objectives: distribution to hospital patients and teachers, to purchase available seed, to clear roads, to assist returnees.’

(UNTAET, 2000: p4–5)

Interventions were also considered for the most part to be timely. A good example of timeliness is the MSF intervention in Uganda, where a rapid assessment was organised one day after cholera first broke out, meetings were held with the relevant ministry, and a cholera treatment centre was established four days after the first reported outbreak (MSF, 1999). Effectiveness was, as ECHO (2000a) notes, at the heart of most individual evaluations. It has proved easier for evaluators to assess an intervention on its own terms than on a wider set of objectives (e.g., impact), even when the latter was required in the terms of reference (see [Chapter 3](#) for further discussion).

When they do consider impact, evaluations raise questions as to the likely long-term impact and connectedness of interventions. Overall, interventions are not considered to have supported capacity building or as likely to support sustainable development. The problem of coordination between different bureaucracies or within one bureaucracy for relief, rehabilitation and development activities has not, for the most part, been overcome.

Furthermore, statements such as those in [Box 2.2](#) are also representative of the broad way in which impact was measured (see also, e.g., ECHO, 1999, 2000a; MSF, 1999).

## Box 2.2 Measures of Impact?

‘What were the results? The assessment concluded that emergency assistance programmes funded by USAID and implemented by U.S. non-governmental agencies clearly helped save lives and alleviate suffering – which, after all, is their overarching objective.’

(USAID, 2000a: p viii)

‘The evaluation team estimated that the intervention directly lowered the morbidity and mortality risk of 250,000 people.’

2000b: 12.9.8)

(DEC,

A minority of the evaluation reports reflected on the appropriateness of using ‘lives saved’ and ‘people fed’ as indicators of impact. This is problematic for a number of reasons:

- We are not told *whose* lives are saved – women’s or men’s, young or old, or from a particular ethnic group or region. Where there is imperfect coverage of affected populations, are some lives sacrificed to save others? Few reports categorise ‘number of lives saved’ within overall excess mortality figures. If there is excess mortality, are lives being saved selectively?
- Like most development indicators, this reflects the priority of agency/government staff. Can more appropriate indicators of impact be developed in association with affected populations? For example, lives saved focuses at the individual level; are there other indicators of impact that focus at the household or community level?
- The indicator ‘lives saved’ tells us nothing about the quality of those lives after they have been saved. As such it is closer to an input than an impact indicator.
- As several of the evaluation reports note, attribution at such an aggregated level is complex.

The literature makes clear that indicator development is central to the process of defining results (Stufflebeam, 1999). As an indicator, ‘lives saved’ may hide more than it reveals and needs to be used with caution if it is to be used as a measure of impact. At the very least evaluators and commissioning organisations should ensure disaggregation by sex, age, ethnicity and region, along with considerations of unintended consequences and categorisation within relevant sociopolitical factors.

Important exceptions to the individual evaluation reports’ narrow focus on effectiveness are the seven DANIDA reports. While noting the effectiveness of DANIDA’s interventions, these also include extensive discussion of the trade-offs between humani-

tarian assistance and political action – to which this chapter cannot do justice. Several of these evaluation reports argue that issues of protection, human rights and humanitarian space are key to measuring impact. This is because humanitarian action needs to be viewed from a wider perspective than the provision of relief and basic services (see [Chapter 3](#) and [Relief, politics and protection](#) for more detailed discussion).

### 2.3.3 Relevance

A major criticism of humanitarian action usually voiced through the media, that relief is often not relevant to needs, is not supported by the evaluation reports reviewed for this chapter. These almost unanimously state that the interventions were relevant, both to the complex emergencies and natural disasters to which they were responding, and to the affected populations. Interventions were not thought likely to significantly disrupt, and in some cases were likely to support, local economies. The provision of shelter and housing is the one important exception to this finding (see below). Environmental damage – caused, for example, by large concentrations of displaced populations and refugees – is covered in only a small number of the evaluation reports and is largely discounted.

However, it is clear that the ‘relevance’ standards used in many of the evaluation reports reflect those of the evaluation team and the agency and government staff interviewed, i.e., not those of the affected population. Because of the apparent lack of adequate consultation with beneficiaries and non-beneficiaries in the majority of the evaluations, it is unclear in many cases whether the affected population considered the intervention relevant or not. However, most of the evaluations that do consider the perspective of the affected populations in adequate detail (for example, DEC, 1999, 2000, 2000b; Norwegian MFA, 1999; Oxfam, 2000) come to the conclusion that interventions were relevant to their needs.

Exceptions include the evaluation of the British government’s assistance provided to Montserrat where the purchase of four second-hand American school buses took on a symbolic value, representing for many Montserratians: ‘an example of DFID incompetence and failure to consult local counterparts in bringing in unsuitable, low-cost equipment that then causes avoidable maintenance problems,’ (DFID Vol 2, 1999: p71). Another exception is the evaluation of ECHO’s programme in Tajikistan (ECHO, 2000i), where the report notes that there could have been greater local purchase and that the most expensive, yet least liked, item in the food baskets distributed was rice. However, views of the affected population also need to be cross-checked, for as DEC comments, perhaps ironically: ‘All beneficiaries and partners responded positively to the use of DEC monies by member agencies, but that is not surprising,’ (DEC, 1999: section 12.9.7).

### 2.3.4 Coherence, coordination and partnership

The issue of coherence relates to individual actors following a common path. It is particularly important where there are many actors, as in humanitarian action. Coordination

is one element of coherence, and deals mainly with operations. Individual evaluation reports do not cover coherence in any detail but focus instead on agency and government coordination. Exceptions include UNTAET (2000b), where the scale of the operation was countrywide.

On the other hand, coherence is considered to differing degrees in the synthesis studies. Synthesis findings are that:

- with the exception of Angola, DANIDA did not have a clearly formulated strategy for assistance in individual protracted emergencies and so lacked visible individual policies for them (DANIDA, 1999a);
- interventions need to pay greater attention to potentially adverse political occurrences (USAID, 2000);
- between 1991 and 1996, ECHO did not manage humanitarian aid in a proactive manner or with a strategic vision (ECHO, 1999);
- between 1996 and 1999 a substantial number of ECHO's interventions prioritised political rather than humanitarian considerations (ECHO, 2000).

Overall, coherence was difficult to achieve, and where achieved may have focused on political ends. Policies and strategies (see also UNTAET, 2000a) for directing interventions were found to be lacking in a majority of cases.

Problems with coordination are generic, and this review endorses the conclusion of the HPN summary report: 'A virtually constant theme in evaluations of humanitarian operations anywhere in the world is poor coordination, along with the retrospective admission that better coordination could have avoided some major problems,' (HPN, 2000: p29; *see also* Spencer, 1998; and Borton & Macrae, 1997). Problems with coordination are noted in about half of the evaluation reports, mainly within and between bureaucracies (for example, within the EC or DANIDA, or between WFP and UNHCR), between head offices and the field, and in relation to the 'tying' of assistance to host-country institutions (for the latter, *see* DANIDA, 1999h).

Coordination at field level is evaluated as being relatively more effective (for example, DANIDA, 1999h). An example is provided by DEC in response to the 1998 Bangladesh floods (DEC, 2000a), where adequate levels of coordination were noted as well as increased coordination in comparison to the 1988 floods. One of the reasons for this is long established and regular cooperation between host country NGOs in Bangladesh that might serve as an example to other countries. Another good practice case comes from Afghanistan and the OCHA mine action programme. DANIDA (1999b) reports tight control and coordination exercised by OCHA over implementing partners since 1989–90, which has provided transparency and accountability and which in turn has meant consistent donor funding. Furthermore, coordination is noted to have helped Afghanistan avoid the difficulties other countries' mine action programmes have faced. One of the main reasons identified for success has been that most of the staff on the programme are Afghani. As the HPN summary report notes: '[coordination] seems an area where some fresh thinking and staff guidance are required,' (HPN, 2000: p29). A focus on good practice may be relevant

here. One important area is the potential role of national and local governments which, as DANIDA notes (1999h; see also DEC, 2000, 2000b; and Van Brabant, 1999), is often overlooked or sidelined by humanitarian agencies during complex emergencies. Another would be to draw on the wider literature on cooperation that has formulated an understanding of why different parties agree to coordinate activities (for example, Ostrom, 1990).

Partnership, as noted by DANIDA (1999a) and ECHO (2000a), is one of the main factors contributing to effective and efficient interventions. The most common partnership arrangement found in the reports is of bilateral and multilateral agencies as funders and UN agencies and INGOs as implementers. This, for the most part, resulted in friction rather than conflict. DEC (2000, 2000a, 2000b) and WFP (1999) demonstrate that agencies in the field did best where they were well established and/or had well established partnerships. USAID similarly found that ‘the fact that the Mission [in Honduras] could rely on experienced partners with a long-standing track record and extensive community networks in areas hit by the hurricane made a crucial difference in facilitating a fast start-up for disaster response activities,’ (USAID, 2000: p36).

### 2.3.5 Participation

Participation, in particular of the affected population, has been noted as a key area for improving humanitarian action and including the affected population as actors. About half of the evaluation reports consider participation by the affected population in humanitarian programmes; findings of three-quarters of this subset are positive, though participation is noted more at the implementation than at the design stage, and is greater in rehabilitation than relief. While this is a common pattern in development cooperation in general, greater attention needs to be paid by agencies to broadening the scope of participation during a humanitarian response.

The idea that ‘community participation’ – defined as different things in different evaluation reports – is likely to lead to more effective interventions is stated rather than demonstrated. For example, DEC comments that ‘community-based needs assessments were highly effective in informing the design of rehabilitation and reconstruction projects,’ (DEC, 2000: executive summary); in addition it notes that hired consultants conducted needs assessments in the first weeks after Mitch, but ‘these outside assessments did not usually involve community participation. The resulting relief package design was in some cases ill-conceived and inadequately conducted,’ (*ibid.*). Other evaluation reports make similar assumptions without providing detailed evidence.

It is important for evaluators and commissioning agencies to establish clear links between participation and effectiveness, with detailed documentation and examples. This will help them to be persuasive in those cases where participation is not considered relevant. In addition, the ‘community’ was rarely disaggregated, so we learn little from the evaluation reports about differential participation by women and men, or young and old, nor whether ‘community’ participation led to more equal shares. Exceptions include DEC (2000) which

covers participation in a gender mainstreamed fashion, and DEC (1999), which considers community as well as individual needs.

### **2.3.6 Coverage**

Most of the evaluation reports consider that both geographical coverage and targeting of ‘communities’ was adequate, with some important exceptions (in particular the ECHO synthesis report, 2000a and UNTAET, 2000a). However, they do not set out clear standards by which coverage was assessed, or was approximate to the DAC guidelines (OECD-DAC, 1999: p23), in this area. We are mostly just told that interventions covered the intended target group, quite often with that group’s participation. But the information missing from most studies is who was not included, who benefited more or less, and levels of diversion.

Detail on coverage appears to be based mainly on agency/government staff interviews and reports. Systematic beneficiary and non-beneficiary interviews may have revealed a quite different pattern. An important exception is the UNTAET evaluation report and its two supporting studies (UNTAET, 2000, 2000a, 2000b). Here the agency self-assessment and external evaluation reports note good levels of coverage; the beneficiary assessment notes, however, that there was a lack of data on the population which led to targeting problems, unequal distribution between villages, inadequate monitoring of the distribution process, and poor quality and quantity of materials. Unfortunately no attempt at triangulation<sup>4</sup> or cross-checking appears to have been made in this case.

ECHO (2000a) comments on the contradiction between ECHO’s stated policy of targeting the most vulnerable groups and the reluctance of ECHO and its partners to exclude less vulnerable groups, even though they do not have the means to come to the aid of the entire population affected by a particular crisis. Given limited agency resources and difficult questions of access, this remains one of the most morally complex areas in humanitarian action.

Needs assessment was generally found to be both participatory and adequate in the individual evaluation reports, with some exceptions (for example, DFID, 1999; ECHO, 2000f; UNTAET, 2000). The synthesis studies, on the other hand, with the exception of USAID (2000a), which did not cover this area, all note that needs assessment should be considerably improved. For example, there does not appear to have been significant improvement in this area during the period of the two ECHO synthesis studies (1991–96 and 1996–99). The evaluation reports that consider communication comment that information provided to the affected population on intervention strategies was inadequate (UNTAET, 2000, 2000a, 2000b; ECHO, 2000d, 2000g; DFID, 1999).

### **2.3.7 Efficiency and cost-effectiveness**

Efficiency measures outputs in relation to inputs to see whether the most efficient approach has been used. Efficiency is mentioned in about three-quarters of the evaluation reports, though only adequately covered in about one-quarter of the set. The question of efficiency focuses mainly on local versus national or international procurement, management issues,

and coordination. Evaluation reports which do consider this in any detail note that agencies did pay attention to the effects of procurement on the local economy and attempted to factor this in to their efficiency considerations (for example, DEC, 1999, 2000b; WFP, 1999). Overall, interventions are considered efficient (for example, among others DANIDA, 1999a; ECHO, 1999, 2000a, 2000e; Oxfam, 2000; WFP, 1999).

Cost-effectiveness is covered in about one quarter of the evaluation reports and, as various studies note (see also OECD-DAC, 1999), proved difficult to measure. Those reports that do consider cost-effectiveness generally note positive findings.

### **2.3.8 Synthesis of cross-cutting findings**

#### ***Use of international standards***

Seventeen of the individual evaluation reports make reference to international standards for humanitarian action, in particular the code of conduct of the International Red Cross and Red Crescent Movement and NGOs in disaster relief, the Sphere humanitarian charter and minimum standards in disaster response, and the EC Madrid Declaration. Most of these references are in passing, with some notable exceptions. This suggests that few evaluators and commissioning agencies are mainstreaming the use of international standards in their evaluations.

This is not surprising, however, since it can be extrapolated that the use by implementing agencies of international standards is very limited. There appears to be a lack of knowledge of the standards, or an inability to put them into operation. UNTAET, for example, comments: 'Initially, it was foreseen that Sphere standards and the Red Cross/NGO code of conduct would be used by all to ensure quality humanitarian response on common standards. However, there has been no monitoring process undertaken along those lines, and there is no evidence that humanitarian participants were reviewing their activities in the light of humanitarian principles. As such it remained essentially a good intention but not an operational procedure,' (UNTAET, 2000: p8). DANIDA (1999f) includes an analysis of making funding conditional on adherence to the Red Cross code of conduct (it considers the Sphere standards too detailed for this purpose).

An example of operational exception concerns MSF-B in Sudan. DANIDA (1999g, Annex IV, Table 2) provides data regarding MSF-B's adherence to international standards of relief programming established by Sphere. While not attaining the output standards (in terms of reduced mortality, for example), MSF-B did adhere to most of the process criteria, and data shows that MSF-B did have an impact in reducing mortality. The report notes that it is regrettable that comparable data is not available from DANIDA's other partners. In addition, ECHO (2000g) reports that the Sphere standards were successfully used by UNHCR in refugee camps in Tanzania.

#### ***Coping strategies***

Coping strategies of the affected population are mainly discussed in the context of development rather than relief, but some terms of reference in the set (in particular the DEC evaluation reports, but also UNTAET) required a focus in this area.

Some evaluation reports provide a, tantalisingly, limited amount of information about coping strategies, but without systematic discussion. Examples of coping strategies from the emergency phase are provided in Box 2.3.

### **Box 2.3 Coping Strategies in the Emergency Phase: Selected Quotations**

‘While this low mortality rate can be seen as an indicator of success in the relief effort, the team’s view was that this low rate of morbidity *after* the cyclone was caused as much by strong local coping mechanisms amongst the people themselves as it was by the initial relief efforts. Beneficiaries emphasised that they had survived after the cyclone by drinking water from coconuts, which were in abundant supply following the destruction of coconut palms.’

(DEC, 2000b: p23)

‘It would be wrong to place international aid at the centre of the rescue and survival activities that followed Hurricane Mitch; instead, neighbourhood solidarity without doubt saved most lives.’

(HPN, 2000: p16)

‘While the emergency response was large and mostly effective, the East Timorese themselves, building on existing formal and informal associations, provided considerable assistance to one another, particular[ly] during the initial emergency period.’

(UNTAET, 2000: p14)

‘The considerable achievement of the people of Montserrat is to have coped with the continuing volcanic threat and then adapt to the devastating effects of the eruption.’

(DFID, 1999: p53)

It is clear from these examples that such strategies are of crucial importance to survival in emergency conditions. Scant attention is also given to longer term coping strategies (for example, migration, household splitting, or patterns of sale and borrowing). The implications of the failure to pay adequate attention to these strategies is noted in DEC (2000b), where an excess of free relief items and lack of clear planning and implementation of the rehabilitation phase for many months after the Orissa cyclone was seen to undermine support of livelihoods. DEC (1999) notes that a belief that ‘famine’ foods were more readily available than was in fact the case led some observers to underestimate the potential for famine.

Other examples of longer term strategies include the response to Hurricane Mitch, where farming households chose to stay put so as to continue farming (rather than, for example, moving to cities). A DEC evaluation comments:

‘The provision of seeds, agricultural inputs and – in some cases – cash, helped farming families remain in their communities, despite massive harvest, soil, housing and livelihood losses. That so many agencies supported agricultural projects constitutes a remarkable and decisive step on the part of the DEC relief effort. Beneficiaries in virtually every community visited by ECATEAMS [multidisciplinary teams of consultants] were overwhelmingly grateful for the agricultural assistance received,’ (DEC, 2000:p13).

Oxfam (2000) also provides examples of the ways in which agricultural programmes built on indigenous forms of mutual support; ECHO (2000g) notes the strong horticultural skills of refugees.

One area of coping strategies hardly touched on concerns psychological and emotional needs. An exception here is DEC’s evaluation (2000), which notes that there was scant evidence that psychological and emotional needs were properly identified and considered in responses. Another important exception is the evaluation report and planning document of the Norwegian Ministry of Foreign Affairs (1999). This thorough and comprehensive document reviews positively overall psychosocial projects in Bosnia-Herzegovina and the Caucasus, after taking into account a lack of beneficiary participation. The report also notes the importance of such programmes in dealing with substantial need in the relief and rehabilitation phases, and how psychosocial interventions can complement more traditional approaches such as food distribution (see also MSF, 1999b).

Clearly there is much to be done by evaluators and commissioning agencies to support understanding of coping strategies in complex emergencies/natural disasters. Particular attention needs to be focused on how those affected manage physically and psychologically, how coping differs by sex, socioeconomic background, ethnicity and region, and how external intervention can build on, or at the very least not interfere with, these strategies.

### ***Linking relief, rehabilitation and development (LRRD)***

The problematic area of LRRD is a recurring theme in the evaluation set, which evidences much discussion, some confusion, and no resolution. Many evaluation reports raise the question of whether there is a linear process from relief to development (for example, DANIDA, 1999g), and also what the appropriate level of intervention should be given the nature of the crisis or disaster. Responses to the latter range from primarily saving lives (for example, ECHO, 2000), to longer term structural change (for example, HPN, 2000; DANIDA, 1999g). Disagreement has been revealed at the most basic level (see Box 2.4).

## Box 2.4 **Conflicting Views of the Links between Relief, Rehabilitation and Development**

‘Addressing urgent emergency requirements whilst taking into account longer-term needs is widely recognised as a difficult, often unsatisfactory aspect of emergency and humanitarian relief responses ... [I]t is difficult, at the level of individual actions, to meet emergency needs in a developmental way.’

(DFID, 1999: p63)

‘Long-term contexts are appropriate to natural hazard response not complex emergencies.’

(DEC, 1999: 12.9.9)

‘Field experience now indicates that a development perspective should be infused into humanitarian assistance efforts from the very beginning of post-conflict aid operations.’

(UNDP, 2000: p37)

‘The notion that relief assistance can be made more developmental in the context of ongoing armed conflicts is problematic ... Nevertheless, emergency assistance programmes can help shape the pattern and direction of subsequent economic development.’

(USAID,

2000a: p ix-x)

Lack of clarity in this area is one of the reasons for decreased performance as interventions move from relief to rehabilitation. The following comment is representative: ‘There is an unclear division of responsibilities between humanitarian aid and development agencies. The expected timeframe for reliance on INGOs in health and other sectors is not widely known,’ (UNTAET, 2000: p16). The DANIDA synthesis report notes that: ‘The quality of intervention was generally high, although there were significant comments about the use of humanitarian budgets to provide a broad range of social services which should have been the responsibility of the local government, as in Angola,’ (DANIDA, 1999a: Executive Summary). Another point concerns the importance of working in partnership with the national government on LRRD. This is made in DANIDA: ‘One clear lesson stands out from the above review. When a transition from conflict to peace is taking place with the full support and participation of the national government institutions and with donors’ backing, there is no real “transition gap”,’ (DANIDA, 1999h: p106).

All of the ECHO evaluation reports reviewed discuss the problematic nature of this ‘grey’ area, mainly in the context of ECHO funds being utilised for rehabilitation and development purposes, a lack of strategic planning, and/or coordination with EC

development bureaucracies. Such problems are certainly not restricted to ECHO, but are generic. There is criticism, for example, of:

- ECHO's support to the housing programme in Rwanda (ECHO, 2000b, 2000c) where moving outside its area of expertise, to housing rehabilitation, caused various problems such as lack of appropriate siting and use of inappropriate materials;
- lack of operation and maintenance training and an exit strategy in Mali and Niger (ECHO, 2000e);
- the creation of dependency of vulnerable populations on state institutions, and of the state on ECHO support, in Cuba (ECHO, 2000f);
- emergency relief substituting for the provision of development assistance in Tanzania (ECHO, 2000d).

It is clearly difficult for interventions to plan for both relief and rehabilitation. An exception, from the Oxfam programme in Colombia, is provided in Box 2.5.

#### **Box 2.5      Linking Relief, Rehabilitation and Reconstruction in Colombia**

The evaluation of Oxfam's programme in Colombia assesses the programme as 'an example of an intervention which consciously sought to combine relief, rehabilitation and development. The design of the humanitarian programme as an integrated response to a multi-faceted humanitarian crisis was key to establishing links with processes of rehabilitation and reconstruction. As with many humanitarian interventions in Colombia, this programme included elements of training and capacity-building which strengthened the communities and enabled them to remain united and firm in their decision to return and rebuild their lives in their homelands in spite of the continuing conflict. In this respect the agricultural component of the programme was absolutely fundamental to the feasibility of this decision.'

(Oxfam, 2000: p37)

This example shows that planners need to be fully aware of the need for an integrated approach should the intervention be likely to have significant impact on long-term development.

The topic is also discussed in the DEC evaluation reports. DEC (2000) makes the interesting but unsubstantiated point that those affected by emergencies do not make a distinction between different kinds of development assistance. Evaluators and commissioning organisations need to do more to promote an understanding of the view of the affected population of how agencies should deal with the LRRD problem. The methodological aspects of this issue are dealt with further in the meta-evaluation section under the heading Implications of the LRRD debate for establishing evaluation standards (4.4).

### ***Relief, politics and protection***

In cases where interventions were focused on natural disasters, or where complex emergencies had taken place in the past and the focus was on rehabilitation, humanitarian space and protection were not generally discussed. Eleven of the individual evaluation reports raise the issue of protection in a substantive fashion, including the seven individual DANIDA evaluations. These also focus on the current uncertainty as to the linkages between humanitarian and political action. Just as many evaluation reports criticise the use of humanitarian intervention as a substitute for development assistance, so the DANIDA reports criticise the use of humanitarian action as a substitute for political action. The DANIDA reports are perhaps better able to cover these issues adequately because they deal with complex emergencies for the most part (i.e., where protection is more of an issue), and because of their broad sweep which covers entire operations, countries, and a comparatively long period: 1992–1998.

The failure to bring adequate diplomatic and political pressure to bear on the parties to a conflict so as to ensure protection is heavily criticised in the DANIDA evaluation reports, particularly in relation to the Great Lakes region, Chechnya, Sudan, and Afghanistan. For example, the thematic report on the UN notes: ‘The massive humanitarian presence in internal conflicts in the 1990s also demonstrated its limited effects on protection. While some humanitarians had a protection mandate (ICRC and UNHCR), unarmed humanitarians could only provide limited protection against intense or targeted violence. UN or UN-authorized military forces sometimes intervened with a humanitarian mandate, but this mostly covered protection of the humanitarian supplies and relief workers – not the intended beneficiaries,’ (DANIDA, 1999h: p4). DANIDA (1999f) notes that inadequate attention to protection stems from poor analysis of conflict, including a focus on socioeconomic issues to the exclusion of political issues. This leads to an inability to design interventions adequately.

As noted in DANIDA (1999h), there is now a greater readiness to consider making humanitarian aid dependant on the safety of the affected population having first been ensured, even though withholding assistance may appear to contradict the short-term humanitarian rationale. This evaluation report suggests that human security – protecting civilian victims of armed conflicts, providing life-saving assistance, negotiating humanitarian access, promoting international and refugee law and human rights – may become the central humanitarian theme during the next decades. It comments that ‘In many conflicts, legal and physical protection is a precondition for survival and hence the possibility of utilising material assistance. In these situations, it is clearly more important for the humanitarians to emphasise protection than to build shelter, or to channel resources for other objectives such as helping returning refugees and IDPs,’ (DANIDA, 1999h: p132).

In the other three evaluations that consider this issue, one notes the need for greater attention to this issue (UNTAET, 2000a), and the moderate performance on the part of agencies (UNTAET, 2000), the second comments on the ways in which lack of

humanitarian space hampered agencies' work (DEC, 1999), and the third records the importance of the presence of Oxfam staff in terms of providing security to the communities with which they were working (Oxfam, 2000)<sup>5</sup>.

### ***Shelter and housing***

About one-third of the interventions supported shelter and housing programmes and as noted earlier, housing was the least successful area of intervention.

Shelter and housing were provided in a variety of forms, from the supply of vouchers worth up to one-third of the cost of the house (USAID, 2000: p31), to houses built under food-for-work programmes (for example, WFP, 1999), to free distribution (for example, UNHCR, 2000), to a mix of all of these approaches (for example, DEC, 2000a). There were also different costs for housing during the same response (for example, DEC, 2000a; WFP, 1999).

While not all shelter and housing programmes were problematic (for example, DEC, 2000 notes, on balance, good results), the following comments are representative: '[P]erhaps the most controversial aspect of the emergency has been the provision of accommodation for those evacuated ... Progress in enabling people to leave temporary shelters has been slow and raises questions about the effectiveness of the emergency housing programme,' (DFID, 1999: p30). Also, 'The shelter kit programme is the most evident shortcoming of the humanitarian response given its delay in implementation and limited coverage,' (UNTAET, 2000: p6).

The main problems included:

- lack of appropriate siting and materials;
- a focus on technical issues, which tended to exclude socioeconomic analysis;
- poor targeting;
- poor management and coordination, including attempts to develop a housing programme within the constraints of an emergency programme. This links to the earlier LRRD discussion.

Recommendations by evaluators responded to these problems, for example, by focusing on house repair (DEC, 2000b), use of local materials (UNHCR, 2000), improved quality control (UNTAET, 2000a) and greater consultation with the affected population (DEC, 2000). The HPN summary report notes that '[Housing] seems to be an area where collective work on documenting experiences and good practices is urgently required,' (HPN, 2000: p28). However, as the problematic nature of the provision of housing has been recognised for some time (for example, RDI, 1990), failure to learn from past experience appears to be a case of lack of institutional memory among implementing agencies.

Issues that are dealt with insufficiently in some of the evaluation reports, even those critical of housing initiatives, are those concerning the appropriateness of housing as well as overall coverage. Evaluators and commissioning organisations need to examine in detail whether provision of housing, which is relatively high cost and focuses on individual

household rather than ‘community’ needs, is a relevant response in the first place. DEC (2000b), for example, notes that building a cyclone shelter in every village was the priority of the affected population. Evaluators and commissioning organisations also need to consider, under the ‘relevance’ and ‘coverage’ headings, if housing programmes selectively cover the affected population, and whether funds could be spread more widely if a different type of intervention was chosen.

### ***Preparedness***

A common finding in the evaluation reports that consider this issue is that governments and agencies are insufficiently prepared. The fact that only two of the evaluations (MSF, 1999b; SCF, 2000) cover an intervention specifically related to preparedness, or with a major preparedness component, suggests that agencies still tend to be reactive rather than proactive in their approach to humanitarian needs, as noted in Borton & Macrae (1997).

DANIDA (1999h) includes an extensive discussion of work on early warning systems in the UN. It argues that: ‘Clearly, there is no shortage of data. As we have seen, by the end of the 1990s databases and analytical sources relevant for early warning purposes had mushroomed. Agency cooperation to share and analyse data was improving. However, the options for action and possibilities for influencing events were limited.’ At the same time, however, this report notes that UN agencies claim that it is difficult to obtain donor funding for contingency planning and preparedness measures. This point is also made by DANIDA which comments that ‘donors continue to be far more responsive to pictures of a full-blown crisis, rather than to genuine early warning,’ (DANIDA, 1999g: Executive Summary). Two of the DEC studies (1999; 2000b) note that a portion of DEC funds should be allocated to preparedness activities. Two good practice examples are provided in Box 2.6.

## Box 2.6 Flood Preparedness in China and Bangladesh

Preparedness was found to be high in the cases of flooding in Bangladesh and China. WFP (2000) reports that the Government of China had contingency plans in place in all relevant departments, including an early warning system for monitoring of water levels in rivers and lakes coupled with an emergency flood diversion plan; evacuation plans for displaced persons; short-term plans for relief and rehabilitation; medium-term plans for dyke reinforcement and resettlement of population; and long-term plans for disaster prevention. The report notes: 'GoC's advanced prevention policy contributed significantly to limiting the scope of the damage from the floods and the loss of life.'

(WFP, 2000: p5)

'In Bangladesh, use of flood shelters built for the 1988 floods was highly effective, and some village communities were so well prepared that they had maps of their locality to indicate the location of the most distressed households, either by type of occupant or by location in relation to high land.'

(DEC, 2000a)

### ***Management issues***

The main management issues raised in the evaluation reports relate to two points dealt with above – coordination and partnership – as well as to implementation procedures, reporting and monitoring and evaluation.

The central management problems raised are described below.

- **Disbursements** (for example, ECHO, 1999; WFP, 2000; DANIDA, 1999g). Evaluation reports note that head offices often lack the flexibility needed in disbursements and, where that flexibility is present, there is often inadequate oversight by donors and reporting by partners on the use of funds.
- **Lack of an adequate rationale for the selection of multilateral or INGO partners.** This finding comes across strongly in the DANIDA and ECHO studies (for example, DANIDA, 1999a; ECHO, 1999, 2000f; also see Borton & Macrae, 1997 for similar findings). The DANIDA report (1999a) recommends that partners should be selected on past achievement and capacity, including adequate monitoring and evaluation systems.
- **Staff shortages, inadequate training and preparation, and rapid rotation of staff** (for example, DANIDA, 1999g). This appears to be a particular problem in humanitarian action where there are, for example, serious problems of institutional amnesia. ECHO draws the conclusion: 'All of the above [factors leading to sound

interventions] imply a need for good staff – the right person, in the right place, at the right time, as one report put it. The evaluations are rich in findings containing implications for partner and ECHO human resources management, including selection, training and evaluation,’ (ECHO, 2000: p7).

Some agencies appear to be responding to evaluator recommendations with plans for institutional strengthening (for example, the ECHO Agenda for Action, ECHO, 2000). The need for competent staff is emphasised by a very large number of evaluation reports in this set and in earlier studies (for example, UNDP, 2000; see also Borton & Macrae, 1997; CIDA, 1997), but appears to have been ignored in many cases.

### ***Monitoring and evaluation***

As one would perhaps expect from evaluation reports, monitoring and evaluation (M&E) is raised most consistently as an organisational theme. Most agencies and partners were found to be weak in this area with considerable consequence. Specific weaknesses included:

- lack of data for analysis and needs assessment;
- lack of clear reporting requirements, with reporting focusing on inputs, and a need for greater emphasis on M&E (for example, among others DANIDA, 1999; DEC, 1999, 2000b; ECHO, 1999, 2000e, 2000h; SCF, 2000; USAID, 2000);
- vague objective statements (ECHO 1999);
- lack of use of logframes (DANIDA, 1999; ECHO, 2000e). Logframes were assessed as being well used in DEC (1999) and Oxfam (2000);
- lack of application of the findings of earlier evaluation reports and lessons learned (ECHO, 1999, 2000c, 2000e).

A large number of the recommendations made in this area relate mainly to strengthening M&E capacity and practice. For needs assessment, USAID recommends establishing ‘a central monitoring and data-collection unit to serve all donors during the early weeks of a complex emergency,’ (USAID, 2000: p58; and see DEC, 2000b). More details on M&E, and recommendations for improving the quality of evaluations, can be found in [Chapter 4](#).

#### **2.3.9 Synthesis of evaluation report recommendations**

The evaluation reports include a large number of recommendations, many of which relate to specific sectoral interventions (food aid, housing, etc). Of the more general recommendations the most common are shown below.

- In relation to coherence, recommendations focus on the need to ensure that interventions are directed by clear policy and strategy. These should deal adequately with issues such as links between peacekeeping, protection and humanitarian assistance; and exit strategies and LRRD.
- Interventions need to be based on an improved level of analysis, in particular analysis of politics, conflict and socioeconomic issues.
- Increased preparedness and early warning need to work hand-in-hand. The case of the 1999 cyclone in Orissa is telling here: affected populations were warned to move because of the likelihood of the disaster, but said that they had nowhere else to go (DEC, 2000b).
- The synthesis reports conclude that needs assessment should be strengthened, including through agencies carrying out joint needs assessment.

- Improved coordination, including coordination with national governments, is recommended by a majority of evaluation reports.
- Better selection and monitoring of partners is a key recommendation in the DANIDA and ECHO evaluation reports.
- In terms of management, the key recommendations relate to the need for an adequate number of competent staff and the importance of overcoming difficulties related to staff rotation.
- Five evaluation reports recommend that greater attention should be paid to gender issues.
- Monitoring and evaluation need to be substantially improved in a majority of interventions.

## 2.4 Meta-evaluation of non-Kosovo Reports

### 2.4.1 Introduction

The central methodological themes in this section are:

- the extent to which evaluation reports of humanitarian action are following current generally accepted good practice;
- the identification of those areas of evaluation reports of humanitarian action in need of strengthening.

These two themes have been the subject of ongoing discussion over the last decade, including within ALNAP<sup>6</sup>. This section attempts to add to the discussion by using the proforma, which is also used in the final chapter for assessment purposes.

Recent meta-evaluations (for example, ECHO, 2000a; UNDP, 2000a) have attempted to introduce user-friendly assessment schemes for the analysis of large numbers of planning and monitoring documents. This approach has been adapted for this Annual Review in order to meet, as requested by ALNAP full members, the objective of beginning a process that would produce agreed-upon criteria for assessing the quality of evaluation reports of humanitarian action. To this end a preliminary proforma has been developed by ALNAP which draws on current guidance and a growing body of what is increasingly acknowledged as good practice in evaluation.

In developing the preliminary proforma a number of good practice sources have been drawn on, including the OECD-DAC *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies* (1999), various academic material, the ALNAP book *Doing Evaluation of International Humanitarian Action: Perspectives from Practitioners* (Wood, et al. 2001) and Apthorpe (2000, subsequently developed as [Chapter 3](#) of this Review). The standards developed by the Joint Committee on Standards for Educational Evaluation (Joint Committee, 1994) were also used (see [Chapter 4](#) of this Review for further discussion of these standards). Greatest attention has been paid to the OECD-DAC *Guidance*, as this is the most comprehensive information available in one source to date.

The ‘quality’ proforma was developed to create a benchmark against which the quality of evaluation reports of humanitarian action could be assessed. It aims to provide a user-friendly flexible tool for intelligent rather than mechanistic application.

The assessment findings of the 33 individual evaluation reports are presented in aggregate form below under the 11 proforma headings. This particular assessment was carried out using ‘adequacy’ as a measure.

The questions of why current gaps and weaknesses exist in evaluation practice, and how these might be strengthened, are considered throughout this section.

#### **2.4.2 Assessment of evaluation reports using the headings**

##### ***Purpose and focus of the evaluation***

As noted in the first part of this chapter, the majority of the evaluation reports have a joint lesson-learning and accountability focus. About half of the evaluation reports note this explicitly. Almost all the evaluations used what might be considered a ‘conventional’ evaluation approach – i.e., they relied on standard techniques, including review of relevant documents, interviews with agency staff and, in about half the evaluation reports, with the affected population. They were also characterised by short-term contract periods, using consultants paid on a daily rate by the commissioning agency.

##### ***Constraints***

In terms of constraints to carrying out evaluations, seven reports note lack of access to specific geographical regions, mainly for security reasons, four note lack of access to data, eight note time constraints, and four note constraints related to the difficulty of tracing and attributing results to funds subsumed within a larger pool of funds not under evaluation. In the latter case, evaluations often assume that where the overall intervention achieved a particular impact, the contribution by the agency being evaluated supported that impact. Care needs to be taken to show this is a plausible assumption.

Discussion of constraints is not extensive. The potential constraint of conducting interviews with affected populations in the presence of agency staff is not commented on in most of the reports, although three of the DEC reports note the importance of interviewing beneficiaries without any agency staff present (see Box 2.7).

## Box 2.7      **Good Practice: Participation, Beneficiaries and Gender**

‘As well as interviewing the agencies’ project officers and key officials in coordinating agencies ... and partner agencies, a sample of beneficiaries will be selected and interviewed by the evaluators. These interviews will be conducted without agency personnel being present. ... The beneficiaries will be questioned on their views of the assistance provided the way they were selected and their overall views of the agency. Interviews with individuals may be complemented by discussions with groups of beneficiaries. So as to assess the agency’s targeting and beneficiary selection methods the evaluation team will also interview a selection of potential beneficiaries who did not receive assistance.

It is expected that the evaluation team will use gender-aware and participatory approaches to seek the views of beneficiaries and, where appropriate non-beneficiaries.’

(DEC, 2000b)

### ***Terms of reference, team composition and time allowed***

The studies provide almost no information about developing terms of reference. It can therefore be presumed that these were issued by the commissioning agencies and generally accepted by the evaluation teams. This is in line with the conventional approach adopted by almost all evaluation reports (see [Purpose and focus of the evaluation](#), [Methodology and transparency](#), and [Opaqueness in evaluation methodology](#)).

Overall, terms of reference were adequate in directing the evaluation to cover key areas of programmatic focus (impact, connectedness, coherence, coverage, coordination, cost-effectiveness, efficiency, and relevance), as set out in the OECD-DAC (1999) *Guidance*. About three-quarters of the reports were organised around coverage of these areas. The OECD-DAC study, and papers and discussion leading to and from it, presumably had some influence in this respect, and there has also been an advance in methodological consistency since an earlier synthesis report (Borton & Macrae, 1997) which noted the lack of standards and commonality in terms of reference.

Terms of reference were less adequate, however, in directing evaluations to cover gender issues and to include representative samples from the affected population. This is probably one of the reasons why limited attention has been paid to these topics. Furthermore, terms of reference did not generally specify the evaluation methodology to be used. While terms of reference may not need to specify the exact methodological approach to be taken, they should at least require adherence to good evaluation practice (for example, transparency in terms of the methodology used, triangulation, and adequate consultation with the affected

population). A good practice case that provides general guidance but does not attempt to micro-manage, providing flexibility for the evaluation team, is given in [Box 2.7](#).

Team composition also followed the conventional pattern, with most teams made up of consultants based in the country or continent of the commissioning agency. While there are potential advantages to this arrangement, including the fact that evaluators may be a 'known quantity' to commissioning organisations, and that the pool of evaluation expertise may be larger in the commissioning countries, there are also disadvantages. These include: the higher costs associated with the use of 'northern' consultants; the limited time they are able to spend on field work (often a result of fee rates and available resources); their frequent lack of appropriate language skills; and the loss of opportunity to build up indigenous capacity. Host country citizens, on the other hand, while not always having the breadth of experience or evaluation background of 'northern' consultants, are likely to have language skills and sound local knowledge and contacts – invariably key factors contributing to an evaluation's success.

As with technical assistance, the choice of team members for evaluations of humanitarian action should complement and strengthen, rather than replace, indigenous capacity (Sobhan, 1997). Of the 33 individual evaluation reports, 21 were carried out by expatriates, 11 by a mix of host-country citizens and expatriates, and only one by host-country citizens (DEC, 2000), the latter specifically to increase host-country capacity. Mixed team reports, i.e. those by host country and expatriate staff, did not appear to be significantly different in quality to those carried out by expatriates only. This combination of team members may currently be the most appropriate team make-up, given that it provides both international and national expertise.

Only one-third of the reports note the expertise of team members. However, this generally only relates to sectoral expertise which makes it difficult to come to any overall conclusions about whether evaluation teams included relevant personnel and skills, including language skills (see section on [Legibility](#) below).

Another feature in nine of the reports (three by NGOs, five by UN agencies, and one by a bilateral donor) was the inclusion of agency staff on the evaluation team, though none discussed the implications of this. While this may increase participation and the likelihood of uptake of recommendations, it may also be seen as a method of biasing results and so lower the credibility of the study. Reports ought to state explicitly the advantages and disadvantages of including agency staff (particularly where the staff come from outside the agency evaluation office) as 'internal' evaluative processes are often viewed as less impartial than those carried out externally. As is well known, negative evaluation results are subject to extensive discussion and negotiation, with evaluation offices often playing the role of mediator between the evaluation team and the department or agency under review. This process itself can lower the credibility of the evaluation because it may appear to call into question the evaluators' impartiality. It is therefore particularly important to provide a clear rationale when including agency staff – particularly as inclusion of agency staff is not a practice that is common in the general evaluation field<sup>7</sup>.

Almost none of the reports note how their evaluation team was selected, whether through tender, head-hunting or, in the case of mixed teams, how internal members of staff were selected. Evaluation reports could add to their credibility if this process and any constraints involved in team selection and composition were noted.

In assessing whether adequate time was allowed for the evaluation, it was presumed that teams would have sufficient time for developing the methodology, reviewing background documentation, consulting with the affected population and agency and government staff, writing and disseminating a draft report, and finalising it. Time for these functions is considered sufficient in 13 of the studies, and insufficient in 15; 5 studies do not note the evaluation's timing.

Overall, consultation with the affected population did not produce adequate information to support impact assessment, or the analysis, in many evaluation reports. For the most part this appears to result from insufficient time being allowed for consultation, and there would seem to be a direct correlation between the quality of an evaluation and the allocation of adequate time to consult with the affected population.

With regard to the allocation of time, there do not appear to be any systematic standards directing commissioning agencies in this area. For example, a team evaluating US\$16m in expenditure was allowed about 200 person days (DEC, 2000); a team evaluating over US\$100m and the establishment of a new state appear to have spent about 120 person days (UNTAET, 2000); and a team evaluating US\$606m appears to have been allowed about 60 person days (WFP, 2000a)<sup>8</sup>. Time allocated may be determined by both commissioning agency requirements and the time that evaluators have available. When expatriate team members cannot spend adequate time in the field, host country team members may have more time available and be able to fit fieldwork around other commitments more easily.

### ***Information on context and intervention***

Provision of contextual information in relation to the humanitarian intervention proved to be one of the stronger areas of the reports – 29 provide it. Analysis of economics and agriculture is stronger than that of geopolitics and social issues, although the latter are mostly adequately covered. Cultural issues, however, are not discussed in any depth, which means that there may have been little understanding of the ways in which culture (for example, intra-household relations, and 'community' or ethnic relations) has affected and shaped humanitarian action.

Only six reports attempt to construct a narrative history. Of these, four include all relevant actors (agency and government staff and the affected population), one includes agency staff and the other includes local level officials. This also reflects the conventional evaluation approach of most of the studies. However, as DAC-OECD notes: 'In attempting to understand situations and structures, to analyse a particular set of events and processes the construction of a narrative history is a powerful tool,' (DAC-OECD, 1999: p18). Specific skills, in particular the ability to listen and sympathise while remaining impartial, and adequate time are necessary for this challenging exercise, especially in discussion with the

affected population. This is thus another area where commissioning agencies need to provide a clearer lead.

Eighteen reports provide adequate reference to secondary reports, and twelve to comparable reports, such as similar studies from another country, region or sector. Finding a balance here was difficult for some evaluators, whose reports read more like PhD theses than evaluation reports, with, for example, extensive historical detail or description of the intervention. Future guidance on evaluation methods should provide details on what can be considered adequate in this area.

### ***Methodology and transparency***

The 33 individual evaluation reports were also assessed in terms of the extent to which their methodology met what is generally accepted to be good practice. Rather than the ‘methodological anarchy’ mentioned in Borton & Macrae (1997), the reports showed considerable consistency in terms of their following what has been termed the ‘conventional’ approach. This approach – that is, the use of different information sources (in this case interviews with agency/government staff, the affected population in some cases, and review of documentation) – might be considered as the mixed-method technique currently promoted in general evaluation theory as a strong evaluation approach. However, few reports cross-checked or made comparisons between these different sources of information. Very few experimented with other techniques, such as self-assessment. Even techniques now commonplace in the development field, such as participatory rural appraisal, were rarely used (exceptions included DEC, 2000 and WFP, 2000; see [Box 2.8](#)). The evaluation of the intervention in East Timor (UNTAET, 2000, 2000a, 2000b) did involve a mix of techniques: a self-assessment, a study of beneficiary participation, and an external evaluation. It thus had an opportunity to triangulate, but this does not appear to have been done.

#### **Box 2.8      Good Practice in Use of Participatory Techniques and Control Groups**

‘Methods included stakeholder workshops, village PRAs, gender analysis, focus group meetings, semi-structured household interviews, and spot visits to FFW sites. Participatory tools included key informant interviews, use of ZOPP-type cards, participatory mapping, calendar making, matrix ranking of FFW activities, and positive/negative impact analysis. Gender issues were mainstreamed throughout the exercise ... To compare the “with” and “without” project situation as well as beneficiaries and non-beneficiaries, mini-PRAs were undertaken in eight severely affected villages, five moderately affected villages and four non-project villages.’

(WFP, 2000)

Almost all reports are based on the concept of the external ‘objective’ evaluation team, and so have not taken on board some of the advances in the evaluation field in the last decade. These include a focus on greater participation of all actors in the evaluation process, greater attention to highlighting good practice in operations, and effective use of evaluation results. In addition, evaluations and their methodological approaches, including those that attempt to remove ‘subjective’ elements, are always influenced by the perceptions and biases of the evaluators. This is clearly reflected in the reports reviewed in terms of their gender blindness and, for example, that some evaluators thought it more important to cover geopolitics or consult with beneficiaries than others. To increase rigour and transparency, evaluators need to make clear how their perspective has influenced their methodological approach (see [Opaqueness of evaluation methodology](#), for a typology that can guide this exercise).

As noted, most reports cover the recommended evaluation areas, with the notable exception of cost-effectiveness. This is covered in only nine of the reports, and while it may be generally felt difficult to capture (OECD-DAC, 1999), where possible an assessment of cost-effectiveness will provide useful information on whether expenditure was adequate. While ‘standard’ areas are usually covered in adequate depth, the quality of the information sources used raises some concern – for the most part these sources are agency staff and agency documents. Commissioning agencies need to make it clearer, when providing terms of reference and guidance to evaluators, that as wide a range of information sources as possible should be used, with triangulation between different sources being an essential methodological element. Not only will use of a wide range of information sources strengthen evaluation methodology, but it will also provide greater credibility, as it adheres to evaluation good practice.

Only one evaluation used a control group approach<sup>9</sup>. This is a striking finding given that quasi-experimental design was required by some of the terms of reference and is increasingly standard practice in development cooperation evaluations. This is one of the weakest methodological areas in the studies and requires attention from commissioning agencies.

### ***Consultation with beneficiaries and the affected population***

OECD-DAC notes that: ‘Experience shows that interviews with beneficiaries can be one of the richest sources of information in evaluations of humanitarian assistance,’ (OECD-DAC, 1999: p25). Despite the considerable emphasis placed on, and experience gained in, this area over the last decade, this set of reports is weak in this area. Only four reports consulted an adequate number of beneficiaries and clearly note the methodology used for this consultation. Thirteen studies did not consult with the affected population. There is only circumstantial evidence of consultation in the remaining fifteen<sup>10</sup>. For example, one evaluation report notes that a decision was taken to consult with beneficiaries, but no further information is provided about this; in others beneficiary comments are dotted throughout the study to support particular statements, but no general methodological details are given.

However, it must be said that evaluation teams found consultation with the affected population easier or more relevant in some situations than others. For example, evaluations that cover several years, such as the DANIDA reports, had difficulty in locating beneficiaries from the start of the period being evaluated. In other instances security of evaluators was a major issue. In most cases, however, it was possible to consult with a sample from the affected population and there does not appear to be any reason why evaluation teams who have done this should not clearly note the methodology they used. In those cases where consultation is already institutionalised (for example, the DEC), commissioning agencies could go further in ensuring that the affected population also plays a role in the evaluation process, and that results are circulated to consulted groups.

### ***Attention to gender and the vulnerable or marginalised***

Only one-third of the reports contained findings on gender, or could be considered to be partly gender mainstreamed. However, even in these reports we learn very little about either the effects of complex emergencies and disasters, or the impact of humanitarian action, on gender relations and gender equality. Reports that do cover gender tended to see gender issues as women's issues, with the result that gender equality was not considered. For example WFP (2000a) quotes a nutrition study which showed that boys in the 6–24 month age group were nearly twice as likely to be physically wasted as girls; the section on 'Gender Issues', however, refers exclusively to women rather than to relations between women and men. The 22 reports that do not cover gender adequately either do not mention gender issues or cover them briefly in a separate section of a few lines.

The reports that do cover gender adequately illustrate that it is an important area and one that it is possible to include in evaluation consultancies. These reports note that women in the affected population play an active role in the relief and rehabilitation phases (a point also made in WFP 2000), and that gender insensitive interventions during these phases can damage longer term efforts aimed at promoting gender equality. In most reports, however, gender is absent, and even basic norms, such as disaggregation of statistics by gender, are not followed. Given that gender, as noted in OECD-DAC (1999), is central to our understanding of complex emergencies and the effectiveness of humanitarian action, it can be deduced that evaluators and commissioning agencies are failing in this area. There may currently be increasing attention to gender issues, for example in the 2001 UN consolidated appeal process (UN, 2001). This appeal process was an attempt by OCHA to integrate gender questions into the CAP, for example, by focusing on women and war.<sup>11</sup>

The reports reviewed fare better in their attention to vulnerable groups in general, with 20 paying adequate attention to groups such as children and the elderly. Once again, however, there is a tendency not to disaggregate and to discuss the 'vulnerable' in general terms. Greater disaggregation would allow a clearer understanding of the intervention's impact.

### ***Coverage of factors potentially influencing performance***

The two areas considered here concern contextual and organisational factors. In terms of the former, about two-fifths of the individual reports can be said to cover adequately the

geo-political context, relations with local authorities and/or the national government, and diversion of assistance away from humanitarian programmes and its potential use by militias and militaries. Given the importance of diversion and corruption in public debates on humanitarian action, commissioning agencies may want to ensure that more systematic attention is paid to this area. Very few studies consider social and cultural factors that might have been a hindrance to intervention impact (for example, intra-household relations or food preferences).

More of the reports cover organisational factors. In particular, assessment of management of the intervention cycle, from design through evaluation, is a strong point of the studies. Implementation and monitoring, for example, are adequately assessed in almost all of the reports. There is, however, somewhat less focus on design and evaluation. Head office/field relations are covered in about one-third of the reports.

### ***Conclusions and recommendations***

This part of the proforma allows assessment: of whether or not key findings have been shared with stakeholders; the logic of flow between a reports' findings and conclusions, and between the conclusions and recommendations; and whether or not the recommendations will be possible to implement .

Twenty-one of the reports reviewed had shared draft versions with stakeholders, usually agency personnel and donors. In the other 12 cases there were either no details on this topic or the draft reports had not been shared. The fact that a majority was shared is encouraging and suggests that consultation on the draft is beginning to become institutionalised, at least at the agency level. However, sharing of findings with the affected population was almost non-existent.

The development of clear conclusions and recommendations, that are linked to the main text and to each other, is a strong area in the reports with just a few exceptions. Evaluators were skilled in summarising their findings and drawing recommendations from these. The level of expression of recommendations varies: 14 studies include recommendations that are specific and include details in relation to implementation; 10 include quite general recommendations, while nine reports either include no recommendations or an unclear set of recommendations. As DAC-OECD (1999) notes, there are no hard and fast rules in the writing of recommendations. Given the diversity of evaluation reports assessed, it was not possible to come to a conclusion as to whether or not evaluators were managing to fit their recommendations to the needs of the commissioning agency or not.

### ***Legibility and dissemination of the final report***

The proforma also allows assessment of the clarity and accessibility of the final report in recognition of the potentially diverse readership, and whether the principal lessons are identified in an accessible fashion.

Overall, 25 of the reports are assessed as clearly written and accessible, and 15 as making good use of maps and diagrams. Twenty-eight reports clearly identify the evaluation's principal lessons in an accessible fashion, and the reports are generally strong in directing different messages to different audiences, through use of the executive summary and lessons learned. In some reports the writing style was unclear, and greater attention to editing would have made for a much stronger evaluation, for example, in the wording of recommendations.

# **CHAPTER 3**

## **EVALUATIONS OF HUMANITARIAN AID IN RESPONSE TO THE 1999 KOSOVO CRISIS: SYNTHESIS AND META-EVALUATION**

### **3.1 Overview of this Chapter**

#### **3.1.1 Introduction**

In June 2000, aware of the opportunity provided by the sizeable emerging cluster of evaluative reports of the humanitarian response to the 1999 Kosovo Crisis, ALNAP commissioned an analytical overview (synthesis and meta-evaluation) of evaluative reports of humanitarian action undertaken within this single context. That original analysis, prepared by Professor Raymond Apthorpe, provided the background paper<sup>1</sup> for the ALNAP October 2000 symposium<sup>2</sup>. This chapter expands on that analysis to reflect issues raised by symposiasts (including ALNAP members) and incorporate evaluative reports published since.

Its twin tasks are to:

1. summarise the principal findings and recommendations by means of synthesis;
2. assess the quality of the reports through meta-evaluation.

Section 3.2, 'The Kosovo conflict: a set of interlocking crises', identifies and reviews problems presented by evaluators as having reached crisis proportions, and therefore their principal foci – whether explicitly stated in terms of reference or introduced by the evaluators themselves. The premise is that the first task of an evaluation or overview of evaluations is to discover how the problem was perceived and review in consequence the appropriateness of indicators used to validate the response.

Section 3.3, 'Agency response issues', selects and synthesises recurring themes in the evaluations' findings, conclusions and recommendations under the broad categories (a) humanitarian principles, practices, and strategic planning, (b) human resource organisation and management, and (c) technical and standards aspects of operations. It seeks to evaluate the strengths and weaknesses of programme design and implementation, based on the premise that evaluation should be multifaceted and consider the wider socioeconomic and political context. The tendency when considering appropriateness of response, for example, is for humanitarian evaluations to focus only on donor-relief supply factors, leaving issues such as agency-driven supply and ideological agendas outside the critical frame.

Section 3.4, 'Meta-evaluation: Kosovo evaluations as learning tools' evaluates the Kosovo evaluations to assess the generic practice of evaluation of humanitarian action, and its strengths and weaknesses as a lesson-learning genre.

Throughout this chapter the term 'programme' is used inclusively, denoting both policy and operational aspects within humanitarian action, as is the phrase

‘organisational/institutional learning’, denoting both the skills and processes associated with knowledge management and implementation.

### 3.1.2 **Characteristics of evaluative reports considered**

The international response to the 1999 Kosovo conflict ranks as one of the largest, in terms of the scale of resources involved, which in turn has generated a multitude of evaluative reports.

The international response to humanitarian needs in Kosovo has been the subject of at least 20 separate evaluations of humanitarian action of which 16 plus 1 synthesis have been published or made available to ALNAP.

While the central findings and recommendations of this chapter are primarily in respect of the ALNAP Evaluative Reports Database (ERD) ‘evaluations’ category, to increase the richness of analysis, it also draws on 50 or so other types of evaluative reports, including ‘After Action Reviews’ conducted by NATO and its member states, parliamentary committees of enquiry, scholarly reviews and seminar reports.

It is the perception of this chapter’s principal author that most categories lack discreteness and most types of report lack serious social analysis. Their management-solutions approach exhibits stakeholder bias that excludes beneficiary perspectives. The sample reports, with the exception of a few French ones, have been made available in English with no evidence of the availability or even existence of other foreign language evaluative reports of the humanitarian response to Kosovo.

A characteristic, particularly in evidence in the ‘evaluations’ category, is the extent to which like evaluated like, with humanitarian agencies engaging ‘ex’ or ‘current’ humanitarian practitioners, and the military, other military.

Most evaluators, although experienced in the undertaking of humanitarian evaluation, had no prior experience of the Balkans.

The majority of the reports focus on single (or umbrella) agencies or cross-cutting themes, but there is no system-wide evaluation and only one joint donor/agency (UNICEF/DFID) evaluation, of limited scope. The lack of a system-wide evaluation, although not uncommon, is unfortunate since what we have as a sample is a collection of uncoordinated and, to an extent, overlapping reports.

As one evaluation notes: ‘Though rarely conducted, joint donor/partner agency evaluations are feasible ... [A] multi-disciplinary team including both “insiders and outsiders” ... can produce a healthy mixture of perspective, knowledge and experience,’ (UNICEF/DFID, 2000: p41). Certainly a series of specialist sectoral analyses, thematic explorations, and so forth, broken down by emergency phase, locale and principal actors, would be invaluable.

The International Crisis Group’s (ICG) *Kosovo Report Card* (ICG, 2000), a review of the international mission’s first fourteen months in Kosovo, although not an evaluation *per se*, comes closest to providing a system-wide overview and yet has a limited focus on

humanitarian issues, is highly selective and, finally, rather abstract despite some illustrative detail in two of its fifty pages.

However, individual agencies are probably still of the view that they can learn more from an evaluation focused on their actions, and the current reality is that there is no appropriate system-wide institution for such evaluations.

Lastly, despite the majority of the reports being on the ERD and within the public domain, several remain restricted. Such practice does not sit well with the ethics of transparency and accountability of non-profit organisations and international agencies.

### **3.1.3 Approach and methods: synthesis & meta-evaluation**

As Roland Barthes put it, ‘the meta-book is the book that talks about the book’. Yet ‘effective synthesis ... [is] not afraid to go beyond the data as necessary,’ (Perrin, *Presentation to the UK Evaluation Society’s Annual Conference*, December 1999). This is the domain of this chapter: ‘proactive[;] ... not afraid to improve, not prove,’ (ibid.).

Every approach to critical analysis has a direct influence on its outcome. As is common in the logic of evaluation theory and practice<sup>3</sup>, the approach adopted here is one that aims to be as inductive as possible, working up from the material being considered rather than down from a deductive, theory-based, schema. Although not mutually exclusive – and particular cases, discourses, paradigms, and comparisons get drawn on by both – the advantage of the inductive is that of not having to force material into pre-conceived grooves, nor having to pass it over if it doesn’t fit. It also allows the adoption of different modes. This paper is somewhat unusual in that ‘aid’ studies tend to be heavily modernist rather than post-modernist, and doctrinal rather than explorative of alternative thinking and positions, in their social science approach.

### **3.1.4 Credibility of the data sources considered**

The value of the synthesis of the conclusions, findings and recommendations presented in this chapter is heavily dependent on the quality of the evaluation reports themselves.

Unlike most of the other evaluative reports considered, the evaluations were undertaken by evaluators experienced in evaluating humanitarian action, and feedback would indicate that the reports were generally well received by their commissioning agencies.

However, despite the fact that the credibility of ‘evaluation of humanitarian action’ as a genre is no longer challenged, the effectiveness of current practice, for which commissioning agencies and evaluators share responsibility, is to be questioned.

The most common approach to evaluation is one of short-term consultancy, often paid for by the agency being evaluated with the objective of making recommendations on how to improve subsequent performance. Unlike social research, management consultancy tends to proceed directly to bottom-line judgements on the issues, often narrowly defined by commissioning agencies in their terms of reference.

Recommendations emerging from such consultancies, although generally well spelt-out, come without qualification or allowance for margin of error. They make no acknowledgement that 'fundamentals forgotten' may be 'mistakes made' or vice versa, and provide no scenario of options with accompanying positives and negatives. Instead, a classical 'single best' business solution is laid out in the 'big idea' tradition of management consultancy.

The majority of evaluations pay scant regard to serious basic description or categorisation, providing broad-brush critiques without in-depth analysis. Where do the militia fit in to the criticisms of military breaches of humanitarian principles? Basic chronologies are uninformed by political economy sensibility or social analysis, and one evaluation report even notes that there had been insufficient time allowed for headquarter enquiry.

The comparative advantage of the evaluation genre may not be for organisational learning (in-house workshops may do this better), and its present standing as a tool for social learning leaves much to be desired. But what can or should one expect from a relatively short-term consultancy by someone with no previous experience or knowledge of the area? More importantly, are evaluations in their current form sufficiently informed to be reliable?

Nevertheless, for all the current problems in respect of practice, quality and resourcing of evaluations of humanitarian action (perceived and real), the story of the humanitarian response to the Kosovo Crisis that emerges here draws particularly on the evaluation genre.

The extent to which lessons learned from previous experience fed through into action in Kosovo remains to be seen. Despite evidence that lesson learning *does* happen: 'The lessons learned in ... Bosnia has [sic] permitted ECHO to avoid a catastrophic outcome [in Kosovo],' (ECHO, 2000o: p9). This is heavily countered by the emphasis in every evaluation that the old lessons have still not been properly learned.

## 3.2 The Kosovo Conflict: a Set of Interlocking Crises

### 3.2.1 Introduction

Every effort must be made to ensure that the emerging story of a complex humanitarian response to a complex emergency is not oversimplified in the telling. The story drawn from these evaluations is multi-faceted, one evaluator even questions whether agencies were responding to the same conflict. There were numerous interpretations and perceived crises, where programmes placed different emphases on the same aspects.

Just as the various parties to a complex emergency understand it from the viewpoint of their own interests and organisation, so it is helpful for this paper to suggest an analytic framework from which to interpret this humanitarian response. Adopting a post-modernist approach allows 'the same' emergency to have 'different' meanings requiring 'different' programmes of action to meet differences in vision, standards, and objectives. This section considers the various crises as perceived by the different actors.

Some reports find it to be a sudden-onset type emergency, the precise, or even approximate, dimensions of which could not reasonably have been anticipated due to scarce, non-existent or illegible early warning signs. The majority takes the diametrically opposite view that its onset was far from sudden and was clearly to be foreseen by anyone looking seriously.

Other reports consider whether this was a conflict that NATO *had to have*, but important as this debate is, the focus of this section is on issues that can be addressed outside it. Given the focus on humanitarian, not military, 'success', this section's concerns extend only to NATO's relatively important role in relief work and whether this role signals an end to humanitarianism as we know it.

### 3.2.2 A crisis of humanitarian principles

What distinguishes this particular humanitarian action from many others is the extent to which it is dominated by the dilemmas and paradoxes thrown up by NATO's involvement – particularly since those governments sending in bombers were also funding humanitarian efforts. Programmes and evaluations alike recognise that NATO's involvement in the overall sequence of events was huge and decisive.

The evaluation reports broadly see NATO's involvement as an actual or potential contradiction of humanitarian principles, and its actions as presaging a new humanitarian order in which cowardly political and diplomatic action on the part of external actors has led to 'humanitarianism by default'. But the evaluations also recognise the contribution made by NATO, directly or indirectly, to the effectiveness of the humanitarian operation (including rapid refugee camp construction, despite some poor siting decisions). One or two evaluations note the military's need for good civil–military/military–civil relations, as well as issues of relative competence and efficiency. The analysis of civil–military

relations is limited primarily to NATO, however, excluding for example INGO–militia relations.

It is fascinating therefore to learn that when the conflict was at its height, discussions concerning humanitarian principles such as neutrality, impartiality and proportionality could, with notable exceptions, be heard more in military than civil circles. As was, and seemingly still is, the case in respect of defining procedures for civil–military relations.

### 3.2.3 A multi-ethnic society in crisis

Like the programmes they evaluate, the evaluation reports perceive the Kosovo conflict as ‘a multi-ethnic society’ in crisis. Only one complains about the distortion and caricaturing by western commentators that reduced the complexities of the 1999 Kosovo crisis to one of a tinderbox of Serbo-Albanian tensions ignited by and for the benefit of Milosevic’s ambition. Virtually none mentions, let alone considers, the different meanings of ethnicity (or multi-ethnicity) as the affected populations themselves understand and act on them.

Like the media commentary, the evaluative reports fixate on ethnicity as for the most part essentialist – primordial *identity*. There are no perspectives on ethnicity understood or interpreted as, say, *image* (that may have been fabricated), or *badge* (that can be pinned on or taken off), or anything else. The reports approach the subject as if all Balkan politics are identity politics, and ethnicity not just one cultural and social marker among others – or, for example, an indicator less of specifically social or cultural distinctiveness than a symbol of geopolitics or a token of history and place.

This weakness with regard to ethnicity and other basic sociocultural, political and economic institutions has serious consequences. Little appears to have been learned from, for instance, Johan Pottier’s work<sup>4</sup> on ethnicity, undertaken following the Rwanda crisis. It makes it difficult for an induction-led synthesis and meta-evaluation to critique the evaluations and other evaluative-type reports effectively on this point, and has major implications for team composition as well as for the writing of terms of reference.

### 3.2.4 A crisis of protection

A critical aspect of this multiple crisis perception is whether programmes and their evaluations saw the emergency primarily as one of assistance or one of protection. ‘Broadly, DEC agencies have been less animated about protection issues in Kosovo than they were during the refugee crisis ... A lack of INGO activity in human rights and protection has (ironically) drawn criticism from one donor, which argues that NGOs’ responses have been largely relief driven and financially driven,’ (DEC Vol. II, 2000c: p108).

Many of the policy review studies and general media reports tend to see crimes against human rights as constituting the fundamental crisis in Kosovo, whereas most of the evaluations, as the programmes they evaluated, focus not on protection but on assistance. This may be partly because at the time of the NATO bombings, when most needed, the humanitarian agencies simply were not there, but it may also reflect perceptions of

mandate. For instance: ‘One reason why NGOs may have been inactive on protection issues in Kosovo is the presence of other organisations who are better able to provide physical protection (KFOR and UNMIK police) or more experienced in monitoring and reporting, and with a special protection mandate,’ (UNHCR, OSCE and ICRC), (DEC Vol. II, 2000c: p108).

For the most part evaluation terms of reference did not specify protection as a point of focus, an omission seemingly unquestioned by evaluators. Should not this uncritical acceptance of terms of reference, and the failure to look at agencies’ broader objectives and policies to identify any protection mandate therein, be a point of concern?

Despite the absence of specific guidance for the evaluation of protection, these evaluations omit it at their peril. If saving life is part of any definition of humanitarian action, it is simply not credible to ignore the issue of protection (including armed protection) in an analytical account of humanitarianism. Whether or not military involvement compromises humanitarian principles or service, the danger of excluding armed protection from the framework and analysis of the humanitarian system is that it will be seen as, and/or become, the sole responsibility of actors with non-humanitarian objectives, priorities and agendas.

It is fair to say that complementary strategies, such as advocacy and affirmative action with protection objectives, were successfully used by some agencies to address the realities of ‘organised criminal violence, gender-based violence, general property rights and the property rights of women in particular’, (DEC Vol. II, 2000c: p107). But only the type of humanitarian protection achievable without the use or threat of armed force was pursued by agencies: ‘During the crisis most [of the DEC-supported agencies] did develop and employ protection measures at different times and in different places, but for some this was not a conscious strategy,’ (DEC Vol. II, 2000c: p101).

Curiously, it is again primarily the military (NATO) reports that address the protection of vulnerable minorities, while the non-military reports tend to make general statements about physical security of persons and groups under threat. This characteristic is particularly noticeable and limiting in any discourse about conflicts of principle and mandate between military and civil actors.

In addition, in a context where human rights violations constitute the crisis, the use of famine and morbidity indicators for humanitarian programmes is highly questionable. The comment that ‘[N]ot a single Kosovar seems to have died from lack of food, shelter or emergency health supports, which provides a very clear indicator of actual achievements,’ (ECHO, 2000q: p5), clearly fails to address the obvious question of how many deaths occurred due to lack of protection.

Overall, the story that emerges is one of (relative) survival *in spite of* the humanitarian effort. One condemnation reads as follows: ‘Agencies’ decisions to withdraw from Kosovo were based on their assessments of security and their ability to continue working. Some point out that NATO prevented humanitarian action from the air, while the Serbs prevented

it from the ground. However by withdrawing en masse, humanitarian agencies, including DEC agencies, effectively failed to sustain “humanitarian space”. In the words of ICRC the “black hole” in Kosovo was “a collective failure of protection”, ’ (DEC Vol. II, 2000c: p106).

### **3.2.5 A relief-supply crisis of spend, spend, spend**

It is not unusual for problems to be attributed to donor-driven, relief-supply factors in humanitarian evaluations. This set of evaluations is no exception. While several note that the overall humanitarian response in Kosovo was driven by geopolitics (pre-dominantly those of the US and UK) as well as NATO’s own agenda, a fundamental problem was that of ‘too much’, not ‘too little’, aid: ‘[O]ne distinct characteristic about the aid effort in Albania over the first three months was the apparent absence of ... any awareness that resources were finite. There was money to do almost anything and to do it almost anywhere,’ (Porter, 1999: p22).

A related, and frequently noted, aspect of the response was its strong bilateral nature with donors bypassing multilaterals to contract responsibilities to their ‘own’ national INGOs. This resulted in ‘uneven standards set by the bilateral actors ... [which] in some camps ... [were] so high that people, only half jokingly, came to ask whether there should be maximum standards as well as minimum ones,’ (Porter, 1999: p22). The majority of INGOs appear to have accepted this lack of impartiality and neutrality with varying degrees of embarrassment, whilst apparently doing little or nothing to reduce or change it.

### **3.2.6 A crisis of international unpreparedness**

A lay-person, knowing little of the humanitarian system, might be forgiven for assuming that any major regional/international emergency, particularly where the big powers had a stake, would benefit from a standing international capacity for rapid emergency response. What emerges from the Kosovo reports is that no such capacity currently exists.

The question therefore arises as to whether our current international humanitarian relief system seriously qualifies as such if it lacks instituted readiness for immediate response, including a capacity for emergency assessment reconnaissance. In the past, this lack has been blamed on cost factors, but material circumstances alone are unlikely to have carried the day. Questions relating to mandates, organisations, rules, regulations and discursive practices must constitute other factors.

### **3.2.7 A crisis of leadership and coordination**

The multi-faceted crisis of the multi-faceted concept of ‘coordination’ is a crisis of the humanitarian system’s own making. The Kosovo evaluations are particularly critical of the humanitarian leadership role within the UN, but it seems inexplicable that the absence of UNOCHA, the UN agency with special mandate for humanitarian coordination, attracts absolutely no attention in the evaluations.

UNHCR, as designated UN Lead Agency, serves throughout as chief whipping boy, with the lowest marks for achievement (for example, due to its ‘missing middle management’),

although higher marks for effort and achievement as events unfolded. WFP gets the top mark for effective deployment of human resources and logistics. UNOCHA, confined to arranging military and civil defence packages from Geneva, producing situation reports in Macedonia, and coordinating with the Albanian government's EMG (Emergency Management Group), was virtually absent, so goes unmarked. UNHCR rather than UNOCHA participated in key EU–NATO coordination meetings.

Lack of strategic planning, and not just information sharing (considered by some evaluations to have been reasonable), carries a share of the blame. This lack was highlighted by Dr Bernard Kouchner (UN Secretary General Special Representative) at an August 1999 meeting in Pristina, when he lamented the fact that there was 'simply no strategic intelligence available' to allow him to tell a visiting new donor (Taiwan) how best to support the relief effort.

The imbalance of participation in Kosovo from within the UN system, as well as contributing to the overall coordination crisis, may reflect a coordination crisis within the UN itself.

A comparison of the imbalance of participation of big UN players with non-UN (but still big) players, including ECHO and OSCE, would be an interesting evaluation focus, as would a focus on UNDP (another part of the UN with notable coordination capacity and experience). It is likely, however, that only a system-wide evaluation would attempt this.

### **3.2.8 A crisis of proliferation**

As with Rwanda, so with Kosovo, what all commentators call 'an unusually large number' of NGOs responded to the crisis. One document famously speaks of this conflict having attracted one NGO for every day of the year. Unfortunately no report goes beyond this rather simplistic 'large number' plaint to tell us just what all these small (and smallest) organisations do, where they come from, what resources they have, etc. They only reveal that having them there makes coordination difficult or nigh on impossible. Their message is of the 'once again' variety where lessons have not been learned.

On the face of it, this is an extraordinarily limited, unconvincing, and trivial assessment. It is most unlikely that fewer of the smaller and smallest would have made that much difference to balanced participation and coordination. It reads less as evaluation than as special pleading.

### **3.2.9 A crisis of unpredicted and unpredictable scale and speed**

Many evaluations share the 'unpredictable nature of the crisis' view. First of all hundreds of thousands of displaced people streamed out of Kosovo and then, equally rapidly and unexpectedly – perhaps especially from Macedonia but also Albania – they streamed back again. To whom was this unexpected?

The evaluations merely report that this was not expected by the international relief agencies. Both the evaluations and the more research-based policy reviews are completely silent on the speed and scale of movement as seen and managed by the affected populations

themselves. This is another instance of the virtual reality of 'foreign-aid land', owing to its social blinkering (and prejudice) about the social, cultural and other institutions of the populations it is meant to be serving.

### **3.2.10 A host nation crisis: maintaining domestic sociopolitical stability**

In this complex emergency, governments were in place in Macedonia and Albania (as of course in Serbia). The Macedonian government in particular considered that the refugee outflow from Kosovo threatened its stability. At the peak of the crisis refugees from the Kosovo conflict constituted 15% of the Macedonian population.

The threat to host nations' domestic sociopolitical stability is, however, only dimly seen and appreciated by the evaluations, where the emphasis in this respect is always on humanitarian and human rights principles and international refugee law issues. Yet, if the Macedonian government had fallen this would undoubtedly have made the whole humanitarian scene even worse.

Instead of treating issues of regional government stability as integral to the humanitarian context, the tendency is to leave the former completely outside the latter's framework, benefiting neither in practical nor policy terms.

What programmes and their evaluations do seem to have been aware of, however, was the regional poverty factor: 'The countries that bore the brunt of the Kosovo crisis are among the poorest in Europe,' (DEC Vol. I, 2000c: p xvii); 'MCIC [Macedonian Centre for International Co-operation] in Macedonia explicitly voiced the need to "balance the refugee crisis with the social crisis"' (DEC Vol. I, 2000c: p 45).

### **3.2.11 A crisis of overbearing humanitarian assistance**

Fortunately, two or three of the evaluative reports make reference to additional problems created by the humanitarian response process itself: '[An] agency complained that the population was fully employed in reconstruction until the international community intervened,' (DEC Vol. II, 2000c: p96). Dispossession is common in foreign-aid land, however well intentioned.

The dangers of negative impact are not unrecognised, but warning voices clearly went unheeded. The following was written in April 1999: 'For a decade, ethnic Albanians have been forced to create their own parallel society, involving every domain from kindergartens to health clinics to architecture schools to a vibrant independent media. It is critical that the international community not "colonise" Kosovo with a sudden "invasion" of well-intended but overbearing humanitarian assistance. Such a misguided deluge of aid last autumn, in the wake of the ill-fated October [1998] ceasefire agreements, seriously jeopardised the integrity and growth of Kosovo's home-grown, still fragile, civil society institutions. It is clear from the many mistakes made by the international community in its efforts to rebuild Bosnia that a peremptory, non consultative approach using "in-and-out" commercial contractors can inhibit the efforts of a vulnerable, war-torn, population to get back on its feet,' (International Crisis Group, 1999a: p27).

### 3.2.12 **A crisis in assistance: assisting a ‘non-crisis’**

Despite all of the above, a theme that appears in some of the evaluations (particularly with reference to assistance) is that ‘no classically defined emergency actually occurred,’ (UNICEF/DFID, 2000: p13); ‘The refugees were relatively better nourished, healthier, and with access to greater resources (including, very significantly, from the Kosovar Albanian Diaspora) compared to those in many emergencies ... [and] mortality and morbidity rates were generally well within emergency levels,’ (ibid.). Also: ‘The humanitarian intervention as a whole can only take very limited credit for the low mortality and morbidity [noted by many commentators]. The quality of the sanitation and environmental health interventions would probably not have been sufficient to prevent large-scale breakdowns of public health, epidemics and, as a result, excess loss of life, had the crisis occurred in a more usual emergency context,’ (DEC Vol. II, 2000c: p25).

WFP’s recognition that ‘coping mechanisms available to many of the affected groups ... played a decisive role in avoiding hunger and malnutrition ... [as] refugees, IDPs, and other vulnerable groups had their own contingency plans, emergency food stocks and resources to overcome periods of dislocation from their homes ... [and] themselves assumed responsibility for their survival and well being, particularly during the first weeks of the emergency,’ (WFP Executive Board, 2000a: p5) should also be carefully noted.

However, no organisation is reported as having considered there was anything but a huge programme of assistance to be undertaken, or that such efforts were inadequate or, in some regards at least, unnecessary.

### 3.2.13 **Calling a conflict a crisis: some discourse analysis**

A final important point is that most of the evaluative reports, particularly the evaluations, tend to speak not of the Kosovo ‘conflict’ but of the Kosovo ‘crisis’ or ‘crises’.

Talk of ‘crisis’ tends mainly to ask ‘what?’ questions (for example, What is the crisis? What is the assistance required?). ‘Conflict’ talk triggers more ‘who?’ questions (for example, Who is responsible for this mess? Who is fighting whom?). Of course there is no watertight divide between ‘crisis’ and ‘conflict’, as the words have overlapping as well as different dimensions.

The word ‘conflict’ makes only very rare appearances – generally where some broader background or scene-setting chronology is done. While the humanitarian response in Kosovo is not charged with fuelling the fire, the evaluations do say that much of the response missed the fire, in respect of protection and coverage. At best, only a third of those affected (those in refugee camps) received assistance.

The discursive practice of calling a conflict a crisis is not however without effect, intended or not. It focuses problem identification on crisis management rather than conflict resolution, peace and development, de-politicising the analysis. This leads to organisations such as the Mother Teresa Society and the Kosovo Liberation Army (KLA), both prominent local players, being in effect written out of the humanitarian scene and its analysis. They are labelled ‘political’ or ‘military’ and hence not ‘approved’ civil-society

organs, which humanitarian agency discourse takes to be apolitical, somehow beyond politics and war. It is a grave gap in evaluation.

Not one of the evaluations addresses its own, or its programme's, mode of discourse. Commissioned evaluation as a genre is rarely self-reflective.

#### **3.2.14 Conclusion: whose crisis anyway?**

The chief findings of this section are that the various actors were responding not to one crisis but to many crises, and that differences in perception as to the nature of these crises led to the use of different success/failure indicators.

The question of whether the crises were found, perceived or made was not asked by either the evaluations or the policy reviews, explaining their tendency to dwell on and over-emphasise donor relief supply-driven factors.

This myriad of crises (perceived or not) will of course have implications for whoever may take on the task of commissioning and planning a system-wide evaluation. It will need to look not only at all the players (international or national) but also at all the crises that lie within this single conflict.

## 3.3 Agency Response Issues

### 3.3.1 Introduction

This section provides some synthesis of the evaluations' principal findings, conclusions and recommendations – with elements of meta-evaluation. They fall into three broad categories:

1. humanitarian policy, practice and strategic planning for emergency response;
2. organisation and management (O&M) issues for emergency response;
3. technical and standards aspects of emergency operations.

### 3.3.2 Humanitarian policy, practice and strategic planning

#### ***To prepare or not to prepare?***

All the evaluations consider preparedness as an issue and most take the view that none of the agencies involved seriously anticipated the scale and speed of the emergency. They find agencies' preparedness capacities greatly in need of improvement, but few make even minimal proposals as to how this might be achieved. There is no examination of how the preparedness lessons should be incorporated into policy, practice and strategic planning.

Given the poor record of early warning systems, it is suggested that investment should be made in the development of rapid reaction rescue capacities: 'The failure of "early warning" in the Kosovo case confirms the historic tendency of such systems to be unreliable or inadequate. Rather than develop its "early warning" capacity, UNHCR should strengthen its mechanisms to react rapidly,' (UNHCR, 2000a: p xii).

#### ***Conflicts of principle: neutral versus impartial assistance?***

It is important to understand the difference between neutrality and impartiality. The principle of 'neutrality' being that of not taking sides in hostilities, and the principle of 'impartiality' that of non-discrimination, a commitment to deliver relief solely on the basis of priority-of-need, regardless of allegiance.

The evaluations of and for INGOs discuss at length whether the military can pursue military and humanitarian objectives that are both independent and simultaneous in a principled humanitarian way. Most of the evaluations conclude that it cannot, seeing a fundamental contradiction in principle. Other reports, particularly those from NATO's lesson-learning workshops, take an opposing view.

Conflicts of principle between military and civil agencies are exemplified by one of the more policy-oriented studies. The issue it addressed most directly was 'whether the Kosovo crisis exemplified the harnessing of the military for humanitarian tasks or the militarisation of humanitarian action, or some combination of both,' (Minear, 2000: p viii). Although many of those involved in the Kosovo crisis, both military and humanitarian contingents, saw this as 'the harnessing of the military for humanitarian tasks. Yes, the harness chafed here and there, but on balance the collaboration was productive. ... we as researchers read

the Kosovo crisis as reflecting the militarisation of humanitarian action, with certain ominous portents for the future particularly in terms of the politicisation of humanitarian access and activities,' (Minear, 2000: p viii).

The differing views are characteristic of the corpus of Kosovo reports, and the issues introduced or intensified by NATO's involvement are seen as vitally important and underlined by all. But while some are afraid it presages 'humanitarianism by default' because of military mandates, objectives and conditionalities, others are more content to emphasise the immediate and temporal specificities, hoping that NATO's engagement brings only limited implications.

Areas of possible conflict of principle and image between military and civil agencies are only part of the picture. Another perspective on the 'conflict of principle' in the Kosovo context is the perception held by many within western society, as well as other cultural traditions, of humanitarian action as a new form of western imperialism.

Finally, clashes of commercial, along with civil and military, principles and practice need to be considered as well: '[T]he humanitarian marketplace. It becomes increasingly competitive, not least within the field of emergency response, and increasingly driven by donors and their priorities,' (DRC, 1999: p20).

### ***Practice ahead of principle: humanitarian evacuation and humanitarian transfer***

The particular pressures of the Kosovo conflict, fast and massive outflows of refugees and resistance to hosting by regional governments combined with media demands, led to the introduction of two new strategies by UNHCR. These were Humanitarian Evacuation Programmes (HEP), involving movement out of the region, and Humanitarian Transfer Programmes (HTP), involving in-region cross-border movement.

These are examined in three or four of the evaluations with, on the whole, similar conclusions. Success in addressing such massive refugee flows would have been unlikely if such innovative practices had not been employed – i.e., principle needed to catch up with practice. 'The innovative HEP resulted in an operation of unprecedented speed and scale that contributed positively to the protection of refugees by alleviating the burden on a reluctant host state that feared destabilisation,' (UNHCR, 2000a: p102).

Problems associated with both are explored including issues of design, execution, vulnerability, eligibility and fairness in selection. '... HTP did not contribute significantly to protection during the emergency. Part of the limited implementation of HTP relates to confusion over the degree of voluntariness or consent required for these refugee movements ... HEP undermined HTP: UNHCR's stated preference for protection options within the region becomes difficult to maintain if evacuations outside the region are available and preferred by the refugees,' (UNHCR, 2000a: p102).

### ***The non-camped refugee population***

A substantial proportion of the total refugee caseload found shelter in what the humanitarian community first called 'host families' and later 'private accommodation'. This latter reference is perhaps not only because payment passed hands but also because non kith-and-kin relations were involved. This is not explored in the reports.

Although they represent the majority, little is revealed about the assistance and protection provided to the Kosovar refugees in Macedonia and Albania that did not go to camps. In Macedonia local religious charities and governments, rather than the international community, probably deserve the principal credit for what was made available. Needs assessment of these non-camped affected populations was even more neglected than that of those in camps, though arguably: 'It was not significantly relevant to the well-being of the majority of the refugee population who were in private accommodation,' (DFID, 2000: p5). However, another notes: 'For the first two months not a single food distribution reached the 27,000 refugees in private accommodation. ... This failure to reach refugees living with host families will be looked upon (alongside the failure to complete the registration of refugees while they were in Albania) as among the gravest shortcomings of the aid effort in Albania,' (Porter, 1999: p22).

The term 'vulnerability' also took on a special sense in this complex emergency. Given the extent to which refugees were in receipt of remittances from relatives living and working, for instance, in Germany, it referred not, for example, to the elderly or the sick, but to those who weren't in receipt of such remittances. There is some evidence that only the poorest went to the camps and that some 'host families' charged rent for the 'private accommodation' they provided. According to Australian media reports the evacuees airlifted to Australia complained that their temporary (Australian) accommodation was far from consistent with what they had had at home.

### ***Coordination***

As mentioned in the previous section, what emerges from the evaluations and other evaluative reports is the complexity of the coordination problem. Individual evaluations tend not to take a comprehensive view but focus only on one aspect or another, making pronouncements accordingly. Furthermore, different stages in the conflict require different assessments. Overall, accounts agree that in certain areas there was little policy coordination to begin with, though this became marginally better as the emergency wore on. UN focal points such as food (WFP) and aspects of education (UNICEF) were apparently very effective throughout and it is also reported that, within and across some more specifically INGO concerns, cooperation occasionally went well.

What varies from one programme or agency to another, in respect of what was well done, may depend on the extent to which each consciously built on earlier experience and evaluations of that experience. The reports however give greater attention to explanations for lack of achievement: 'For reasons that are well-known, the Kosovo crisis brought in very substantial funds from member states and each wished to maintain a high profile

in-country, including undertaking their own programmes. This led to considerable confusion and lack of coordinated activities,' (ECHO, 2000k, 2000l, 2000m: p5). But note also that: 'Within a few weeks as donor liaison mechanisms also gained in effectiveness, a more coherent and consistent needs-driven approach was evident. Coordination, nevertheless, proved particularly difficult to handle in this emergency. ... We should have articulated the many problems and constraints more clearly from the start,' (Morris, 1999: p17).

What may come as a greater surprise is the comment that, despite shortcomings, adverse effects can be exaggerated: 'UNHCR shortcomings [in assistance and coordination functions] ... did not have grave consequences for the welfare of the refugees: indeed they were relatively minor in relation to the overall relief response. ... [Nonetheless] areas of demonstrated weakness and inability to rapidly meet its own standards of response affected the credibility of the agency [which therefore could be considered to have suffered the most]', (UNHCR, 2000a: p xi).

While the evaluations mention the constraints of consensual coordination, none make recommendations for change in this area nor examine it in other ways.

### ***What constitutes an emergency?***

As with any disaster or humanitarian crisis that occurs in a sovereign state, a state of emergency cannot be acted on until it is declared and recognised as such by the national government concerned. The simple question of whether there is an emergency to meet may be part of the reason for little or no response at the outset. It would also appear that key institutions lack clear rules and guidelines for consideration of what constitutes an emergency status and how it should be acted on.

Where guidelines exist, to what extent are they known and used? 'The underuse in UNICEF (even ignorance) of the emergency handbook throughout the operation is notable,' (UNICEF/DFID, 2000: p34). The same evaluation comments: 'Software may not be totally emergency friendly,' (ibid.: p34).

In terms of whether a situation merits being regarded as having emergency status or not, one evaluation reads as follows: 'Between April and May 1999, morbidity and mortality rates in Albania and Macedonia remained below the threshold that signifies an emergency. ... [F]ar below the [mortality] threshold of greater than 1[death]/10,000[people]/day, the key Sphere indicator for an emergency. ... Data for this period are not available from Kosovo or the rest of the Balkans. Since the return of the refugees to Kosovo a comprehensive health study undertaken among the Kosovar population indicates similar low levels,' (DEC Vol. II, 2000c: p21).

There are also different types, scales, and durations of emergencies, each perhaps with their different requirements. With UNHCR the problem was partly that it 'responded to the Kosovo refugee crisis as if it were a "normal" emergency. Standard routines for a smaller

or slower emergency were followed (although not always attained),’ (UNHCR, 2000a: p x).

It is to their credit that some of the sources draw attention to these important considerations. Training issues that need to be informed by them are mentioned under the heading ‘Human resource weaknesses’ (see [3.3.3](#) below).

A further area of concern is the use of indicators. Normally in evaluation a different concern is considered to deserve a different indicator, as is the case in most of the Kosovo reports. However, to gauge the success (or failure) of a programme in terms of, say, a ‘no one died of famine or disease’ claim, when neither famine nor disease constituted the emergency is, as some evaluations say, downright misleading, if not aberrant.

### **3.3.3 Organisation and management of human resources**

#### ***Lack of standing rapid response capacity***

As mentioned earlier, another striking story to emerge is that of an international humanitarian community that lacks a rapid response capacity. Rescue here does not refer to the humanitarian evacuation programme that eventually got underway in Kosovo, but to a strategic concept for initial agency reconnaissance. The US government’s DART (Disaster Assistance Response Team) approach, which emphasises the assessment required for such rescue, may come close to what the international humanitarian system lacks.

Most evaluative reports, including those by the military, lament this woefully inadequate situation. For example: ‘Recommendation 1: If ACT is to be operational [in emergencies] it must have an emergency management team that it can deploy immediately to make preliminary assessments and to prepare programmes. Where a multi-agency response is involved, this should include preparing common programming,’ (ACT, 2000: p10).

The lack of a specialist standby rapid response, whatever may account for it, is an absurdity. If one is ever established, particular attention should be paid to the rapid assessment skills required and whose capacities are to be built – those of ‘foreigners’, or of ‘nationals’. What are needed in the first instance are rapid response assessment teams, not rosters of active-service personnel. One evaluation notes that emergency response capacities have yet to be mainstreamed into UNICEF.

Although the above is illustrative of the wider picture, it is encouraging to note that in response to a recommendation put forward by the evaluation WFP commissioned, the Emergency Preparedness and Response Steering Committee was created at executive staff level to ‘establish a better framework for early warning, preparedness and response to rapidly evolving emergencies,’ (WFP Executive Board, 2000b: p1).

UNHCR’s evaluation noted that UNHCR was ‘not established to provide relief in emergencies ... If members of the General Assembly want to do this, they will get a refugee agency that is quite different from that they established 50 years ago.’ It noted, however, that: ‘It is clearly possible to turn UNHCR into a superbly efficient rescue service,’ (UNHCR, 2000a: p xiv), if and when this is the policy decided.

### ***Training and briefing***

Several evaluations comment on the paucity of available training material and consider this to be partly responsible for poor practice. This is another example of a crisis of the agencies' own making. Happily some evaluation terms of reference called for an assessment of personnel orientation, guidance and training before and during deployment, but the evaluations generally paint a very gloomy picture. When does poor practice qualify as malpractice?

In terms of briefing: '[S]ome pointed out that if they had been briefed about field work and the importance of cooperation ... by a veteran field worker, the message might have been more convincing than coming from a desk worker,' (DRC, 1999: p25). Debriefing is also crucial, as is the use of the information gathered from the undertaking.

Agencies were noted as having recognised the useful combination of in-house training with external courses by organisations such as RedR, Sphere, and Merlin, but the reports, despite their vocal criticism, offered little remedy other than one interesting suggestion that a travelling contingent of field-craft trainers should be established.

The symposium background paper, on which this chapter is based, proposed something akin to the 'learning support office'<sup>5</sup> currently being developed by ALNAP.

### ***Human resource weaknesses***

Poor programme performance can often be attributed to poor human resource management (HRM). Unprofessional employment and people management appears to have been much as usual in this complex emergency. A familiar pattern emerges in respect of short-term assignments, lack of appropriate training, even lack of briefing on arrival. Gender is reported to have been mostly 'forgotten again', and the high level of staff turnover a destabilising factor, particularly where agencies lack systems and procedures to ensure continuity and enable new staff to become effective more rapidly.

Shortcomings are also identified in relations between foreign and national staff; between country offices and points of delivery; in the mismatch of specialists and generalists (mostly the absence of the former where required sectorally); between routine and volunteer personnel; and in respect of conditions of service and duration of deployment/assignment. One evaluation notes that: '[P]eople who have been deemed unfit for previous operations in the Balkans appeared again in the Kosovo operation,' (DRC, 1999: p25).

Weaknesses in the management of organisations, especially middle management, are easy to diagnose. A common finding in many of the evaluations is that: 'In massive emergencies the agency should ensure the rapid release of middle managers by the immediate adoption of directive, rather than voluntary, deployment practice,' (UNHCR, 2000a: p xvi).

These weaknesses are responsible for many of the undesirable actions and outcomes in the programmes performed. However, other than in particular cases (primarily the O&M

reports), little provision is made either in the terms of reference or the evaluation team composition to allow organisational and managerial dimensions to be competently addressed. This is probably why there continues to be so little progress in the evaluation of these areas, even though the evaluations make a few more proposals on the remedy for this than they do for training and briefing.

All the evaluations portray Kosovo as yet another case where a large-scale humanitarian response pressured agencies into recruiting and deploying personnel without training, although UNHCR personnel did report that they had received a certain amount of appropriate training.

The problem seems to lie particularly within INGOs, where even senior staff members were found to be unaware of their organisation's commitments as signatories to codes that affirm, among other things, certain human resource management norms: 'Although 5 of the 12 DEC agencies are signed up to the pilot People in Aid "Code of Best Practice" some senior staff in DEC agencies were unsure whether their agency was a signatory or not. It is not possible to say how well agencies conformed to the code,' (DEC Vol. I, 2000c: p xxvii).

### ***The importance of existing presence and national staff***

It appears to be the exception rather than the rule that agencies insist on existing presence and national staff as a condition of response. However, in addition to improved timeliness and contextual understanding, it is suggested that agencies working through local partners are less likely to pursue their own agendas.

Two evaluations clearly note existing presence and national staff as enhancing factors: '[T]he effectiveness of this church work is due to a combination of factors: 1) they already operate on the ground ... ; 2) they had staff in place (whether national or missionary) who spoke the local language ...' (Tearfund, 2000: p27); and in the case of UNICEF: 'The predominance of national staff in middle management positions had a direct impact, in that it facilitated a deeper understanding of needs and realities in the three countries visited than is the case for international agencies managed almost uniquely by internationals, most of whom do not speak the relevant languages,' (UNICEF/DFID, 2000: p5).

### ***Lack of social knowledge and learning***

Another important aspect of the story to emerge from most evaluations is the 'once again' factor. Once again 'foreign' humanitarian workers arrived ill-equipped in terms of their sociopolitical and cultural knowledge of the conflict-affected populations. Despite the code of conduct that the big INGOs have all signed up to (introducing an ethical dimension), social and cultural ignorance of beneficiaries was the norm. The heavily loaded term 'beneficiaries', as discourse analysis would say, accounts for a large part of the problem.

An obvious example is the criticised UNICEF 'Superman Project' developed as part of a mine-awareness campaign. As well as conveying the wrong message – that Superman will save you – when the intention was aimed at inculcating safer behaviour: 'Handicap

International psychologists recommend the identification of role models within the community whose behaviour children can copy, rather than the use of fictional heroes ... [T]he age range of the target audience (7–14 years) is too wide, there are no cluster bombs featured although cluster bombs form part of the threat, and the material is in any case not immediately available in Serbian,' (UNICEF/DFID, 2000: p24). This highly illustrative example is typical of many from other agencies.

Since it is well recognised that people in the face of a crisis deploy a range of coping strategies, including the decision to move, why did the programmes and their evaluations discover so little about these strategies? One factor is that results-based management is neither oriented nor sympathetic to social knowledge needs, another is the genuine difficulty of systematic consultation in the first phase of an emergency response. This is, however, no excuse for a lack of consultation in evaluations that come later with more time.

The majority of evaluations do too little to cross the knowledge divide but the very nature of complex emergencies is that they are just that, complex: 'Consultations and cooperation with local communities are part of DRC [Danish Refugee Council] project practice. But emergencies, it was argued, left no time for consultation other than those with local authorities. The principle, all agreed, was valid. But the question was how [to do so] in the heat of the emergency,' (DRC, 1999: p22). Agencies' basic day-to-day work inevitably involves adjustments and learning, clearly not undertaken in a social vacuum.

Even where evaluation time is dedicated to learning about beneficiaries, through in-depth individual and family interviews, the focus is mainly on discovering whether assistance provided was appropriate rather than learning, for example, why a return 'more spontaneous than organised, and self-resourced than assisted,' happened (UNICEF/DFID, 2000: p32).

### ***Lack of organisational learning***

The question of organisational learning relates to much of the above. Like most humanitarian evaluations before them, the Kosovo reports state that past supposedly learned lessons still have not been learned (or at least where learned have not been acted on). One evaluation records a member of staff asking: 'Why are questions about contingency planning and related staffing policy only now discussed. Since many years WFP has been handling emergency operations all around the world, and WFP's budget has shifted from the original 80% for development to almost 80% for emergency programmes. Why are constraints like these raised only now?' (WFP, 2000b: p27).

The fact that the evaluations make many explicit comparisons with earlier complex emergencies might help speed along the information base required for such learning. However, where previous evaluations are the source of such wisdom, use of them should include critical – and transparent – examination of their methodology.

While the need to focus seriously on the process of organisational learning is not being addressed, there are signs of more robust attempts at lesson-learning from Kosovo. Evaluation terms of reference and team composition should be revisited to acknowledge the linkage between lesson-learning and accountability, rather than adopting the view that addressing both represents overload.

### ***The UN Lead Agency role***

UNHCR was selected as UN Lead Agency in Kosovo due to its refugee protection mandate under international law, although it is not a humanitarian relief organisation by mandate or, arguably, capacity. Why ‘refugee crisis’ should trump ‘food crisis’ or ‘integrated and system management crisis’ ( and, by extension, WFP and UNDP) is not explored, but the evaluations do note a lack of conceptual clarity around the ‘Lead Agency’ status, and its voluntary (consensual) nature. None however discusses it seriously nor makes recommendations.

Accreditation, particularly of the large numbers of NGOs, appears not to be part of the process of participation. Although raised as an issue, no evaluation offers recommendations, and overall the evaluations pay insufficient attention to the Lead Agency as an issue.

### ***Delivery, distribution and impact***

All the reports agree that, for whatever reason (they never address why): ‘Uniform tracking, monitoring and evaluation systems and approaches were not applied in the emergency,’ (UNICEF/DFID, 2000: p7). Other similar statements: ‘Neither the ICRC nor the Federation has a monitoring system that provides sufficient feedback,’ (ICRC/IFRC, 2000: p ii); ‘[O]f particular concern is the absence of a system which reconciles logistics statistics (what was sent out) with relief statistics (what was distributed to beneficiaries),’ (ibid.: p iii).

According to one of the military reviews: ‘The US ambassador in Albania did not have a good understanding of the [Marine Air Ground Task Force’s humanitarian] assessment team’s role. ... The consequent lack of coordination and cooperation resulted in the departure of the assessment team before it had completed its mission,’ (US Department of Defence, 2000: p106). Although UNHCR’s Village Needs Assessment Unit in Pristina was seemingly much appreciated it is not assessed in any of the evaluations.

One reason why programmes do not monitor their work in this area may be that agencies still live in the charity paradigm and are simply unwilling to dedicate funds to this purpose. This sits particularly badly in the case of well or over-funded programmes.

Most evaluations that mention ‘needs assessment’ conclude that remarkably little was done. The extent to which, in this emergency as in so many others, relief supply-side considerations and blueprint approaches ruled, was also noted. This was exemplified by responses that at one point threatened a mini-crisis in assistance because they were so inappropriate to needs: ‘Unsolicited “supply driven” in-kind donations clogged the system

(second-hand clothes). In some cases they contravened standards and guidelines for emergency aid (milk formula),’ (UNICEF/DFID, 2000: p12).

The following appears typical of the scene as a whole: ‘Uniform tracking, monitoring and evaluation systems and approaches were not applied in the emergency, nor was adequate capacity available for these functions,’ (UNICEF/DFID, 2000: p7).

Similarly: ‘There is an absence of meaningful reporting. Copious information – like the volume of inputs distributed – is available on the result of projects, but [there is] a paucity about the impact on the beneficiaries along with their families and communities, limiting to a great extent possibilities for more pro-active planning in projects,’ (ECHO, 2000k, 2000l, 2000m: p4).

As usual, cultural factors as they relate to needs appear to have been ignored, despite cultural sensitivity supposedly having been a factor in some recruitment. For example, most of the detailed accounts of psychosocial assistance observe that this to have been the area of service most dogged by definitional and conceptual problems, as well as by culturally inappropriate ‘Western’ ways of diagnosis and treatment. However: ‘The psychosocial programme was appreciated by a large number of recipients (“*Very useful and we needed It,*” was a typical comment), even though psychosocial support was relatively slow in coming as it was treated as part of health and not as emergency relief,’ (ICRC/IFRC, 2000: pi).

It appears that food needs in Kosovo were over-emphasised, for example, in comparison to shelter requirements, and winter needs were identified too strongly with shelter needs.

This synthesis takes the view that an unmonitored programme, besides being unworthy, is, in certain respects ‘unevaluable’, particularly now that standard evaluation of humanitarian action criteria – e.g., appropriateness, effectiveness, efficiency, connectedness, coverage, complementarity and coherence – feature widely in evaluation terms of reference.

### **3.3.4 Technical and standards aspects of emergency operations**

#### ***The Sphere Technical Standards***

Despite the Sphere Technical Standards becoming more widely known and discussed, with supporters seeking to expand coverage beyond the current five technical sectors, the evaluation reports (and ALNAP Kosovo Symposium plenary discussions) indicate that they remain controversial, both in respect of their value in principle and their applicability in practice, where operational flexibilities and strategic priorities must carry the day. The principal findings of the evaluations as a whole are summed up in the following: ‘The evaluation team found that awareness and application ... of the Sphere standards was poor within most DEC agencies, their international networks and local partners. ... Some DEC agency staff questioned the relevance of [the] Sphere standards in a European setting ... [H]owever, the evaluation team felt that most of the Sphere standards were applicable, particularly those relating to assessment, monitoring and information systems,’ (DEC Vol. I, 2000c: p xxvii).

### ***Codes of conduct and operating principles***

As already noted, evaluations found little evidence that codes of conduct were familiar to the majority of agency or implementing partner personnel: 'The evaluation team found that awareness of the code of conduct and its principles ... was poor and little attention given to their application. ... Many local partners had never heard of either the code or Sphere standards, and where DEC agencies worked through local or international networks, other agencies in the network did not necessarily share the same commitment to standards. There was very little monitoring of adherence to the Code and standards in agencies' planning and implementation of programmes and no requirement from the DEC to report against them,' (DEC Vol. I, 2000c: p75).

### **3.3.5 Conclusion: saving graces**

The findings, conclusions and recommendations above amount to what can only be called a severe critique of the aspects of the programmes they address. It is a damning indictment of the humanitarian system that the findings and recommendations synthesised here are neither new, nor Kosovo-specific.

That the less defensible aspects of the international response (including failure to reach the majority of those presumably in need) did not have a greater negative impact owes more to luck than judgement. It deserves to be remembered that, in this instance, needs were reported as apparently lower than typical in complex emergencies.

The expression 'saving grace', used in one evaluation, is taken up in another: 'The independent evaluation of UNHCR performance during the Kosovo refugee crisis acknowledges that the two main "saving graces" were the hosting by families and the refugees' ability to pay for rent and food. ... It also notes that this situation was unsustainable. In other words, the third "saving grace" was the short duration of the crisis,' (DEC Vol. II, 2000c: p67).

This synthesis adds three saving graces of its own: the stability of the governments in the region; NATO's humanitarian contribution (though NATO is not normally considered part of the international humanitarian system); and the strength of public opinion in terms of not wanting 'another Bosnia'.

Almost without exception, the evaluations fail to explain what is arguably the most prominent social feature of this conflict and its response. That is, how the affected populations themselves saw and managed both the conflict and the vast and rapid movements that constituted their response to it. As well as elements of spontaneity and reflex, there must have been some institutional, structural and social capacity factors.

This synthesis cannot treat the issue substantively due to the absence of social information and analysis in the reports considered. As a result, it is hard to reconcile, for example: 'The generous mentioned provision of more or less timely and adequate food to the target

groups by the WFP and other major food providers,' (WFP, 2000b: p20); 'mortality and morbidity rates were generally well within emergency levels,' (UNICEF/DFID, 2000: p13); and, 'the finding that the majority of refugees and internally displaced persons were in fact well nourished meant that there was little need to distribute the high protein biscuits,' (ibid.: p19).

That a good part of the story is, anyway, one of a human rights crisis turned into a humanitarian crisis is more to the point.

## 3.4 Meta-evaluation: Kosovo Evaluations as Learning Tools

### 3.4.1 Introduction

This last section is an exercise in meta-evaluation of the broader enterprise of evaluation of humanitarian action itself, as revealed by the reports. The aim is to determine and then deconstruct some aspects of its nature and scope as currently practised, in order to look for the comparative advantage, and disadvantage, of such an evaluation as a learning tool. The purpose is to come up with, and consider proposals for reform.

Evaluation's dual objectives, lesson-learning and accountability, may make it a less effective learning tool than activities dedicated solely to learning – e.g., lesson-learning studies or learning workshops. If so, where ought the emphasis to lie in terms of a reasonable expectation of the commissioned evaluation type? This chapter shares the concern expressed in three or four of the Kosovo studies that evaluation is perhaps currently asked to do too much, and forced to grapple with too many types of expectation, for which insufficient provision is made.

Part of the problem with present practice is that evaluations are more in line with what could be called knowledge-enhancing practices than with contributions specifically to knowledge management and implementation. As lesson-learning practice, it is not critical enough – indeed in some regards distinctly *uncritical*. This can be traced to the fact that evaluation is often restricted in its focus to the stated policies and objectives of the programmes being evaluated. Assuming these to be sound, it proceeds to verify their attainment.

Other limitations stem from the lack of serious social and political analysis apparent throughout. This leads to remarkably simplistic ideas about treating the affected populations in terms of having 'ethnic problems' only. As a result, the social, cultural and political local context of programmes almost never receives the quality and quantity of attention it ought, even in those reports that at least attempt to give it attention.

Such limitations are also clearly a manifestation of the constricting timeframes imposed on the consultancy framework. The adoption into humanitarian discourse of the phrase 'humanitarian crisis' instead of 'Kosovo conflict' has insulated the evaluative reports from even 'ethno-nationalism', a term that appears only once or twice in the 50 or so sources. There is uncritical and unanalytical reference to ethnicity, as if this were all that social and political analysis of the local and national context required.

The paucity of training is an important factor emerging as part of the story told by the evaluations, but is also an issue for evaluators in respect of the skill composition within evaluation teams.

### 3.4.2 Humanitarian agencies, actors and acts

All the Kosovo evaluations agree that evaluations should focus on more than just operations (and agencies). However, they are often discouraged or prevented from doing so by their commissioning organisations.

What an organisation is and what it does is never solely defined by mandate. For one thing, few such mandates are descriptive, rather, they are exhortatory. For another, organisations have staff, resources, histories, problems, strategic objectives and so on, each an influencing factor in what an organisation is and does. This meta-evaluation takes the view that other levels of analysis are equally important, in particular the analysis of individual actors and their acts.

As shown earlier, individual actors (agency personnel) may not be familiar with, let alone observant of, the codes to which their organisation subscribes, or even their organisation's mandates. A focus on acts allows for a comprehensive assessment, whatever the mandate and ethical code of an organisation. Even where mandates and codes are known, how they are understood and interpreted may be more influential to what is said and done than what a mandate says and intends. The evaluative reports show that beliefs vary considerably, even within the same organisation. Saying and signing is not necessarily believing and practising. The importance of circumstance and situation is above all not to be written-off in emergencies – hence the importance of the post-modern approach in this chapter.

Humanitarian organisations consider themselves to be non-commercial but nevertheless perform certain commercial acts, such as surviving as organisations through marketing and public relations. Similarly, commercial organisations (in this instance the military) have non-commercial acts to perform. In other words, not everything a relief organisation imagines, says, believes, and does, is relief and not everything the military imagine, say, believe, or do, is concerned with combat. There are multiple realities that determine what a particular organisation does in a particular situation, which may throw up concerns about the degree of 'prostitution' of humanitarianism. A debate about just humanitarian versus non-humanitarian organisations (i.e., the military) misses out too much to be practical or intellectually acceptable.

In sum, it is important for evaluations of humanitarian action not to accept stated objectives, policies and principles as givens. A more thorough and nuanced approach will be able to account for the realities of practice in relation to principle.

### 3.4.3 Policy evaluation contrasted with project evaluation

Evaluation of humanitarian action tends to lean heavily towards project evaluation, this being the primary reference of the increasingly standard criteria: appropriateness, efficiency, impact, coverage, connectedness and coherence (yet to be appraised for their value with regard to the evaluation of humanitarian policy).

Unlike previous evaluations, the Kosovo set of reports is notable for the extent to which the reports combine a policy and project/operations focus, undoubtedly due to the policy issues thrown up by NATO's involvement.

However, little more than introductory considerations on policy analysis were included in the 1999 OECD *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies* (pp 24–25). A key requirement of the evaluation profession will be to determine what the evaluation of project and policy have in common, as well as their differences.

While both policy and project (i.e., programme) evaluation requires recourse to objectives and indicators, these are not always provided by the agency project or programme under evaluation. The WFP evaluation remarks that: '[T]he difficulties in evaluating [WFP's response] can be found in the lack of specific objectives and assumptions formulated, as well as in the lack of proper indicators to measure achievement and assess the assumptions,' (WFP, 2000b: p2). A frustration echoed by others.

Induction-led evaluation expects to find the criteria it needs for its own purposes provided, as it were, on a plate. Where this is not the case, if evaluators cannot find ways to devise and provide these for themselves, they are exposed to the charge of not getting their job done. The responsibility, however, must be shared by the agency being evaluated.

#### **3.4.4 Lesson-learning and lessons to be learned**

Lesson-learning is stated as a principal objective of most of the Kosovo evaluations (also most of the policy review documents and workshop reports). They return again and again to the fact that fundamental lessons for strategic planning, the organisation and management of the humanitarian response, and indeed its monitoring and evaluation – which may well have been recommended by previous evaluations – remain to be learned.

Most of the Kosovo evaluations consider that some of the shortcomings of the humanitarian response are due to *force majeure*. The circumstances of this conflict are seen to be exceptional in various ways, and therefore requiring special allowance to be made. This meta-evaluation has no general quarrel with special pleading, provided what is considered to be specific and perhaps excusable is clearly distinguished from what is not. However, what it does seek to ask is whether, in part, typical evaluation practice is as much to blame for lack of lesson-learning as are the evaluated organisations themselves.

Shortcomings in present and past evaluation practice are certainly part of the total picture to be explored. Terms of reference are typically overloaded. Divisions of labour in evaluation teams tend to be on a sectoral or functional basis only. Conclusions and recommendations of the 'single best' form – i.e., not menus of options with the advantages and disadvantages of each option spelt out – is the dominant practice. But on top of this comes the handicapping trend of seeing lesson-learning as simply filling knowledge deficits (a practice all too familiar in evaluation of economic development cooperation), uncomplemented by serious attention to knowledge management and implementation.

Current evaluation practice, team composition and so forth, is singularly ill-designed and ill-equipped to address the lesson-learning purpose that it sets, and has had set for it, unless it can shift its focus to organisational (institutional) learning. This learning is usually a task

that agencies set for themselves, but to which evaluation could make a useful contribution. Unless the evaluation process recognises its role in relation to institutional learning, it will lose its status.

### 3.4.5 **Opaqueness in evaluation methodology**

Lack of methodological rigour results in impressionistic findings that are hard to defend, and undermine the reputations of humanitarian evaluations. Giving sufficient attention to methods and methodology is critical to both the quality of, and confidence in, evaluations.

Methodology (M) is sometimes neglected as an assumed 'given', understood by all. In reality there is no common approach or understanding, and denial of any use of methodology in the undertaking of an evaluation is not unknown. It is critical to recognise methodology's multiple-layered nature. The following four layers are therefore suggested: M-1, M-2, M-3 and M-4.

M-1 is the layer dealt with by most of the Kosovo reports. This includes the timetable, who was where, when and why during the mission, the basis for the division of labour in the team and the time allocation in the field and other locations. Although extremely limited in analytical value, it reveals the basics of how an evaluation proceeded and whether it spent enough time in the field and HQ.

The M-2 layer should include policy analysis methods, evaluators' involvement in finalising the terms of reference, team management issues, established networks, and how difficult relations in the field were addressed – whether with representatives of governmental and other agencies, or with those of the client organisation.

The M-3 layer should cover conceptual and analytical or methodological frameworks for the study, including precise field methods for data collection, the criteria against which field information was assessed and judgements reached. It should also outline the extent to which evaluators felt the methodology worked and implications for findings, conclusions and recommendations.

Clarity of criteria is particularly important where the influence of team members, as optimists or pessimists, can be both subjective and influential – i.e., is it assessed as a positive that a programme achieved so much or as a negative that it failed to achieve everything. Such assessments should not be totally subjective.

Only one evaluation mentions methodology in this way: 'Our assessment is qualitative. While a method of marking that awards points on a scale, of say 1 to 10, may be appropriate for a simple project evaluation, it is much too simple for evaluating agency operations in a complex historical context. ... The standard tools used are those of historical and public policy analysis,' (UNHCR, 2000a: p3). But information about which tools and what they yielded, or not, compared with others, is not given, nor can it be deduced from the text of this example.

The final, M-4 layer, ought to guide the critical reader as to how the evaluators drew conclusions from findings and recommendations from conclusions. These reports do not. Given that the client (among others) is likely to judge recommendations not so much in terms of 'true or false' but rather 'feasible or costly', a methodology that is transparent would tell us the parameters used to reach such recommendations.

Despite the associated difficulties, it is all the more important in the face of complex field situations, multiple actors, lack of data, variable translation of field interviews, limited access to the field and beneficiaries, and difficulties with attribution, to give prior attention to methods and methodologies. However, the flexibility needed to respond to issues and ideas emerging from the fieldwork, and to cope with resistance met in the field, must be maintained. A two-stage fieldwork approach allowing an initial field visit or tour of the study area can help focus the development of appropriate methods and methodologies.

Short-term consultancy is unlikely ever to compare favourably with research, committees of enquiry, select committees, independent commissions and so on, in terms of its hard data findings. But evaluations are normally expected to yield *some* hard findings. One evaluation claims to believe, but unfortunately fails to demonstrate, that: '[W]ith regard to final results, the most important finding of the evaluation is the high level of satisfaction of beneficiaries,' (ICRC/IFRC, 2000: p i). This level of aspiration is exceptional in evaluation, being more typical of research.

Greater transparency in addition to less overloaded terms of reference is needed. Everyone ought to develop frames and discourses that permit greater focus on conclusions and recommendations, including the provision of a menu of different scenario options in the recommendation section.

The comparative advantage of evaluation as a lesson-learning tool needs to be reconsidered from time to time. It is the functions, purposes and methods of evaluation, and not the genre itself, that is at issue – including the extent to which evaluation should be linked in practice with other approaches to learning (and participation and ownership).

#### **3.4.6 Independent evaluation: achieved, necessary, relevant?**

To what extent is the independence of the evaluation really necessary? And what does it seek to achieve? Independence may well not have been achieved even where evaluations claim that status.

Is independence more strategic than other considerations for credibility – such as expertise, authority, or sometimes nationality? Does it function more as a 'credibility-placebo' where it is no guarantor, in and of itself, of quality of output?

The reality of independence is that it can be tenuous, particularly where the like-by-like approach to commissioning extends to evaluators carrying out commissioned evaluations for past employers.

To deconstruct ‘independent’ further, and ask independent from what? – the configuration of particular disciplines, ideologies, sectors etc. should be considered, as well as personal values and premises. To whom does the evaluator report? An agency’s top management or board, a beneficiary union, an ombudsman, or another sort of watchdog? These issues are not necessarily unrelated to the choice of one consultant over another.

That the stakeholders that are taken seriously are not generally beneficiaries or their representatives, or the public that contributed money to appeals, is another important part of this picture. The stakeholder might find token reflection in evaluation team composition and the M1 features of methodology, but when even this tokenism is missing, greater concern arises in respect of what is done in the name of independent evaluation.

This chapter’s principal author believes that:

- mixed teams of insiders and outsiders are more likely to provide credible analysis and recommendations (most Kosovo evaluations fall into the category of outsiders only); and,
- in some circumstances, invited consultancy has a better chance of being read and heard by the client than an unsolicited paper in a professional journal.

A USAID ‘evaluation tip’ states: ‘If objectivity and credibility are key requirements, an external evaluation may be the appropriate choice, whereas if stakeholders’ ownership and acting on it are priorities, more collaborative or participatory approaches are usually better,’ (USAID CDIE, 1997: p3).

Given that complex emergencies inevitably involve many agencies and that interagency relations are likely to be an important factor in any humanitarian response, more interagency evaluations should be undertaken with their potential for increasing transparency – possibly a more critical characteristic than token independence.

#### 3.4.7 **Paucity of social, cultural and political analysis**

Applying a humanitarian label should not obscure the political causes or consequences of a complex emergency. According to a recent *Disasters* journal article: ‘It is now part of received wisdom that humanitarian assistance in conflict and post-conflict situations may be ineffective or even counterproductive in the absence of an informed understanding of the broader political context in which so-called “complex political emergencies” occur,’ (Cliffe & Luckham, 2000). The reality of the majority of the Kosovo evaluations is a poverty of social, cultural and political analysis.

This virtual void has huge ramifications in terms of what is learned and not learned about the conflict and conflict-affected populations. Economic analysis of the cost-effectiveness of operations is for the most part also missing, despite the fact that regional economic poverty is widely considered to have had an impact on every aspect of the response – including dilemmas for relief exit strategies.

Is this analytical *unintelligence* simply due to a lack of appropriate skills on the part of evaluators and programme managers? In addition to increased training, the growth of a

rights-based approach to relief might have the invigorating effect of forcing evaluators to consider local people, their perspectives and contexts.

A final area of consideration, supported by some of the evaluations, are factors influencing the effectiveness of faith-based agencies' relief work as compared with non-faith based ones. A number of key dimensions should emerge from such a comparative study, though some might relate more, for example, to the issue of existing and perhaps continuing presence and national staff than to faith and ideology.

#### **3.4.8 Programme outputs, outcomes and impacts**

As could be expected from evaluations that, for the most part, display no social learning with regard to the affected populations, they offer little by way of impact analysis. Some note this themselves. Others proceed as if anecdotal evidence were enough. On the whole the evaluations do not meet their terms of reference in relation to social learning and impact assessment. However, the terms of reference that call for social learning, but do not make sufficient provision for its undertaking – in terms of time, resources and team composition – are equally to blame.

There also remains the problem of the extent to which evaluators press on regardless to their conclusions and recommendations, demonstrating a lack of self-awareness and an unwillingness to challenge the constraints placed upon them.

Spelling out conceptually the different requirements for outputs, outcomes and impact analysis, as compared with problem identification and inputs, might help remedy this situation. Here again social, economic, cultural and political intelligence and outcomes and impact analysis are closely related.

#### **3.4.9 Attribution of evaluator bias**

If the concept 'independent' is difficult to pin down, 'bias' may be even more so – especially as this could be considered to be ever present in all social and other enquiry. The issue for meta-evaluation is how best to detect and demonstrate it, so that allowance may be made for it.

One of the co-authors of the 1999 DANIDA synthesis report, *'Evaluation of Danish Humanitarian Assistance 1992–98: Volume 1,'* remarked on the extent to which he was able to trace 'evident bias'. Although the seven evaluations shared the same terms of reference and looked at aspects of the same donor programme (albeit in different countries), the six evaluation teams came from different institutions, with different histories of evaluation and composed of different sectoral bases. These characteristics shaped frames and discourses to the extent that if team names were covered up, anyone with knowledge of the evaluation community could have guessed which evaluator had carried out which evaluation.

However, the Kosovo evaluations did not have common terms of reference, and, though looking at the same conflict, were not all concerned with the same programme. It is

therefore not immediately apparent how a meta-evaluation that was heavily induction-led might identify independent variables or go about the rest of the task of attributing bias.

As for bias in respect of this meta-evaluation, there is no reason to believe it is any less influenced by this author's beliefs, experiences and formation as a social anthropologist. These include strong assumptions that commissioned consultancy and freely undertaken research deliver very different products, varieties of interdisciplinary scope, verifiability of findings and conclusions. This in addition to the view that evaluations of humanitarian action are best met by outcomes-oriented work, as is the case for evaluations of economic development issues.

#### 3.4.10 Conclusion

The task of meta-evaluation is not easy, not always welcome, and often not understood. It is not easy because of, among other things, the frequently apples and pears nature of the evaluation reports reviewed. It is not always popular because evaluators and their managers may see it as a threat to the validity of not just exemplars of a genre but the genre itself.

Evaluation has been widely advocated as a path to lesson-learning (as well as accountability), and despite the critique in this paper, the contribution to humanitarian knowledge and analysis that the Kosovo evaluations represent is itself a testament to what present practice can achieve.

The Kosovo reports are diverse. While some may be described as continuations of management by other means, others (some are mixed of course) suggest that evaluation might best serve lesson-learning through advocacy, verification, and witness (particularly where protection is their main concern), rather than through would-be management mandates, business and organisational solutions.

A standard evaluation thesaurus sums up what are particularly evident as tools of inductive argument and critique as: 'analogies, examples, counter examples, counter explanations, more often than ... exact rules ... and the statements it use[s] ... are rough guides to the truth, that is, hints and heuristics rather than exact laws. We use certain modifiers – like *prima facie*, balance of proof, and *ceteris paribus* – sometimes *probably* – to flag the qualifications involved,' (Scriven, 1991: p221).

Whether it would be well for different types of evaluation to proceed differently remains to be examined. With regard to lesson-learning, if the aim of evaluation is to do more than investigate mandates in order at least to meet the needs of results-based management concerns, it will have to find ways of becoming more outcomes-oriented than it is at present. This in turn means evaluation managers and evaluators becoming less socially blinkered or blind.

That more and more evaluations are being done and increasingly placed in the public domain – and given public consideration as in this publication – is, on any reading, excellent. What has not been explored in this chapter, however, is how the actual

commissioning, designing and doing of evaluation could best be institutionalised in order for it to meet everything that is expected of it.

# CHAPTER 4

## OVERVIEW OF THE ANALYSIS AND NEXT STEPS

### 4.1 Introduction

This chapter combines the analysis of the principal findings and meta-evaluation contained in [Chapter 2](#) and [Chapter 3](#), thus covering forty-nine individual evaluation reports, five synthesis reports and one summary report.

In synthesising the principal findings from both the Kosovo and non-Kosovo sets, section 4.2 highlights the differences, considers likely contributing factors, and the implications of the overall findings for the humanitarian system and for ALNAP's own agenda.

As well as combining the meta-evaluation from the two preceding chapters, section 4.3 adds to the analysis of report quality through an application of the preliminary proforma to the Kosovo set. It concludes with a consideration of the next steps needed to strengthen the role of evaluation within the humanitarian system and, again, the implications for ALNAP.

#### 4.1.1 Contextual differences between the sets

The two sets of reports differ significantly in terms of the contexts in which the operations being evaluated took place. Of the complex emergency/natural disaster contexts covered by the non-Kosovo set only East Timor displays similarities to Kosovo, in respect of the efforts made towards peace enforcement, the setting up of an interim administration, nation-building, etc. Only one East Timor evaluation was available for inclusion in this Annual Review.

Detailed analysis of per capita expenditures in each of the operations evaluated was not readily available to the authors. However, perhaps the most striking difference between the Kosovo and non-Kosovo sets is the sheer scale of resources deployed in Kosovo, far outstripping those deployed in the other contexts (see [4.2.1 below](#)).

Despite such obvious differences between the contexts, many of the Kosovo evaluations are comparable to the non-Kosovo evaluations in that they generally shared the same criteria, were generally undertaken by consultants, and focused on one particular area or sector.

### 4.2 Principal Findings & Issues Raised by the Reports

#### 4.2.1 Directions of funding

The pronounced bias towards Kosovo in the allocation of funds by the humanitarian system has been raised by other publications (e.g., IASC, 2000), and discussed by one report as follows:

‘While donor governments gave US\$207 per person through the 1999 UN appeal for Kosovo, those in Sierra Leone received US\$16, and those in the Democratic Republic of Congo little over US\$8 ... The cost of Camp Hope in Albania for 3,000 refugees for two months was equivalent to the 1999 UN appeal for Angola ... In Africa for the past 20 years, refugees have often had to survive on lower than basic food requirements. Kosovo refugees in some cases were receiving Turkish delight in complimentary food packages. In the emergency, 80,000 refugees were flown to other countries in Europe, something that would never be conceived of in an African crisis,’ (DEC, 2000c vol 1: p77).

Such disparities clearly show that the international humanitarian system is strongly influenced by factors such as national interest, geopolitical strategising and media focus. As a system it falls significantly short of meeting the criteria necessary for the use of the term ‘humanitarian’, such as impartiality, independence and considerations of proportionality, as contained in International Humanitarian Law and the Red Cross/NGO code of conduct.

A necessary first step in equipping the system to counter such biases, is the monitoring of humanitarian expenditure on a systematic basis. The Inter-Agency Standing Committee’s publication *Global Humanitarian Assistance 2000* (IASC, 2000) represents a welcome development in this regard. It was obliged, however, to make use of datasets that were not always comparable or consistent, and it is not clear whether it will be produced on a regular basis. Efforts need to be made by the DAC, OCHA and the system generally to strengthen the datasets essential to such monitoring.

A second step would be work to develop measures of humanitarian need to enable agencies involved in funding and coordinating a response to determine at an earlier stage than at present whether the humanitarian needs of affected populations (however defined) were being met, not met or significantly exceeded. Initial work in this field is planned by the Overseas Development Institute and Tufts University. Although it will inevitably encounter conceptual and technical difficulties, it can build on accepted existing approaches in the nutrition, food security, and public health sectors, and make full use of the widely recognised Sphere Minimum Standards in Disaster Response (Sphere, 2000).

Another funding-related issue emerging from the reports is that a significant percentage of the funding being evaluated, perhaps between 30 and 50 per cent, concerned rehabilitation activities rather than activities which addressed ‘acute human suffering’. These proportions are based on an impressionistic assessment of the activities evaluated and are very approximate. Nevertheless, they raise important questions as to the overall appropriateness of impact indicators such as ‘number of lives saved’ for an assessment of the overall intervention. In addition they point to the need for datasets on humanitarian expenditure to make a clearer differentiation between relief and rehabilitation.

#### **4.2.2 The results of humanitarian action**

Judged on their own terms, interventions were assessed by the evaluations as being largely successful at meeting requirements and their stated objectives, with inputs leading to

appropriate outputs. This in itself is a remarkable achievement. The efforts of agency personnel, which appear to be one of the most important factors contributing to success, need to be highly commended, given the context and their difficult working conditions. At the same time, the claims of success need to be qualified.

As [Chapter 2](#) and [Chapter 3](#) point out, the stated objectives of interventions were often vague. The measures by which these objectives were assessed – mouths fed and lives saved – could have been more focused and disaggregated. In particular the Kosovo set tends to focus more on organisational issues and coordination or, at a lower level, on the results chain, either inputs or outputs. We learn for example that funds were used to rebuild housing, but not the uses of that housing (Tearfund, 2000).

Evaluation of impact – that is the real difference the activity has made (for a full definition see OECD-DAC, 1999) – was required in many terms of reference, but commissioning agencies appear to have been satisfied with reporting on effectiveness or input instead. The challenge in this case is promoting an understanding of the difference between effectiveness and impact, and ensuring that both areas are adequately addressed in evaluations.

Interventions were also assessed by the evaluations as being for the most part relevant to the affected populations in both the Kosovo and non-Kosovo set. In this case the standard of relevance does not appear to include systematic assessment of the views of the affected population, undermining overall findings.

In addition, the Kosovo and non-Kosovo evaluation reports were consistently weak in key methodological areas (notably triangulation, consultation with the affected population, and lack of attention to gender). They were for the most part opaque in methodology, to a much greater extent than is the norm in evaluations of development cooperation or the general evaluation field.

Few of the reports raised the issue of attribution adequately and only one attempted counter-factual analysis. The Kosovo reports in particular were unable to isolate the main causes of results – e.g., what was due to the good health and coping strategies of refugees, and the support provided by host families, and what to humanitarian action. Determining the factors responsible for a particular result is a major function of evaluation, and so this inability to isolate the main factors represents a significant shortcoming. This failure to establish credible indicators of success, combined with weak methodological approaches, must call into question the validity of the conclusions reached by many of the reports.

#### **4.2.3 Coverage**

The majority of evaluation reports concluded that coverage of affected populations and vulnerable groups within the programmes evaluated was at the least adequate, and that resources were not in general diverted. That resources were provided to vulnerable groups is a significant achievement and should help counter criticism in this area. On the other hand, both the Kosovo and non-Kosovo sets illustrate that the issue of coverage needs to be

accorded more careful consideration by evaluations. In the Kosovo case a number of reports point out that the major focus of interventions in Albania and Macedonia was the refugee camp populations, to the exclusion of attention to refugees in host families. In the case of the non-Kosovo reports, the lack of attention to IDPs is similarly noted. Another issue relating to coverage is that while assistance may for the most part be reaching vulnerable groups, this does not mean that all of the vulnerable population is being reached, a point not adequately considered in a majority of the reports.

#### 4.2.4 Lack of attention to protection

A common theme from the evaluation reports is that humanitarian agencies are not giving adequate attention to protection, humanitarian space and human rights in complex emergencies. This point comes out most clearly in the seven DANIDA evaluations which note to differing degrees the importance of protection as a key strategy. It is common for agencies to focus on the provision of material assistance and services such as health, food and shelter, leaving protection to agencies such as UNHCR and ICRC that have specific protection mandates. While UNHCR (2000a) notes the creditable performance of that agency in providing protection to refugees in the countries of asylum around Kosovo, [Chapter 3](#) argues that lack of protection, and particularly armed protection of vulnerable minorities by the humanitarian system and the international community within Kosovo, should have been a main focus of those evaluations given the human-rights violations there.

Within the humanitarian system there is now greater recognition than before that all humanitarian agencies have a role to play in relation to protection. A range of activities is underway which should contribute to an improvement in the situation. For instance the ICRC Protection Workshop Process, begun in 1996, has held four workshops so far (e.g., Von Flüe and de Maio, 1999), and IASC will shortly publish a practitioner's field practice manual (Paul, 2001).

However, debate continues as to what constitutes an appropriate role for different types of humanitarian agency. Paul (1999) notes that in considering their role in relation to protection, NGOs are often concerned with how their involvement in protection activities might affect their ability to carry out their mission; jeopardise their principle of neutrality; and risk their access to areas or countries controlled by groups/authorities antagonised by their protection activities. Addressing the issues raised by the evaluations is likely to require genuinely collective system-wide efforts to establish common understandings and approaches.

In addition, there is a parallel need to develop standards and criteria for use in the evaluation of protection. For instance UNHCR (2000a) notes that UNHCR was criticised for being both too timid and too forceful in its relations with the Government of the Former Yugoslav Republic of Macedonia and in respect of the initial refusal of asylum to refugees. Evaluations can support the development of such standards, and the UNHCR evaluation (2000a: [Chapter 6](#)) is itself an example of good practice in the use of standards for

assessment of protection (even if limited to refugees from Kosovo and not human rights violations in Kosovo).<sup>1</sup>

#### 4.2.5 Coordination

Coordination has a well-established reputation for being amongst the most problematic performance areas within the humanitarian system. That both the Kosovo and non-Kosovo evaluations support this view comes as no great surprise. Kosovo provides one of the most telling examples, where UNOCHA was not called on to play a main coordination role and UNHCR, designated as the UN Lead Agency, was to an extent marginalised by NATO and bilateral donors for what were largely political reasons (see [Chapter 3](#) and UNHCR 2000a). There are also many other examples in the non-Kosovo evaluations (see [Chapter 2](#)).

Debates on the most appropriate modalities for coordination in the system continue, with little advance on those that began in the 1970s. While some praise and support OCHA, others criticise and undermine it. Likewise, some argue strongly in favour of the Lead Agency model, while others argue equally strongly against it. The adage that ‘all agencies want coordination but none want to be coordinated’ remains useful in explaining the coordination problem. However, the extraordinary persistence of the problem can probably only be explained in terms of structural factors such as the large number of separately funded, separately governed agencies typically involved in any single operation. Short of radical measures to reduce the number of agencies or channel funds through a limited number of central coordinating bodies, the system’s structure will probably only achieve significantly improved coordination if it is able to develop an appropriate framework of incentives and disincentives for agencies to coordinate. Currently, the rewards for good coordination and the penalties for not coordinating are just too weak to have a noticeable effect.

The persistence of the problem also raises issues for evaluation. Are evaluations dealing adequately with coordination as a central problem in the system? Almost certainly not, as their primary focus is on the performance of single agencies with coordination frequently regarded as a secondary issue. This argues for more collaborative evaluations. Are the necessary coordination standards and indicators in place for use by evaluators? Again the answer is negative, and this lack needs to be urgently addressed.

However, the picture provided by the evaluations on coordination is not all gloom. Examples of good practice are revealed in, for instance: OCHA’s Mine Action programme in Afghanistan (DANIDA, 1999b); actors in the 1998 Bangladesh floods (DEC, 2000a); and the EMG in Albania (DFID, 2000). A common factor in each is a significant host country input – from Afghani staff in the Mine Action programme, to Bangladeshi government and NGO input, to the location of the EMG in the Prime Minister’s Office in Albania. This relates to the need for good partners on the ground, a point agreed by all the reports that covered the issue. Partners were usually conceptualised in terms of INGOs, NGOs or multilateral institutions, but can clearly also include different levels of national government.

#### 4.2.6 Linking relief, rehabilitation and development

While the non-Kosovo set of evaluation reports raises LRRD as another of the most difficult areas, half of the Kosovo evaluations do not. Their focus is on the relatively brief period when the main refugee population was outside Kosovo. However, the seven ECHO Kosovo evaluations find that LRRD remains a major concern for ECHO, and the main points raised in the non-Kosovo reports are very similar to those in the Kosovo reports that do cover the issue:

- a lack of policy and strategy to guide LRRD;
- a lack of communication within and between agencies;
- an inability to plan adequate exit strategies.

These results are consistent with a major recent study on overall (rather than just humanitarian) assistance to post-conflict recovery:

‘Despite ostensible good intentions, too often aid promised has not been committed, aid committed has not been delivered, and aid delivered has arrived too late. Moreover, the planning and implementation of reconstruction aid has frequently suffered from inadequate preparation, poor coordination and lack of perseverance,’ (Forman and Patrick, 2000, p1).

The so-called ‘grey-zone’ between relief and development is not just a problem for the agencies being evaluated, it is also a problem for evaluation. Given that a significant proportion of humanitarian action budgets are being expended on both rehabilitation and development activities, evaluation reports need to be more specific in their recommendations concerning LRRD. An example of good practice in this regard is ECHO (2000p) that recommends a timeframe for hand-over; indication of the names of agencies intended to take over work; and a mid-term review of the hand-over process. A useful exercise would be the collection and dissemination of good practice in the evaluation of LRRD by ALNAP. Ideally this would be undertaken in collaboration with the Conflict Prevention and Post-Conflict Reconstruction Network.<sup>2</sup>

#### 4.2.7 Coping strategies

There is evidence from the reports that the issue of coping strategies is on the agenda of some agencies, even if it is still dealt with in a sporadic fashion. While the Kosovo and non-Kosovo complex emergencies/natural disasters were very different events, affected populations used similar kinds of coping strategies. Primary among these strategies was mutual support. In the non-Kosovo cases [Chapter 2, Box 2.3](#) has already referred to the importance of neighbourhood solidarity in response to Hurricane Mitch, and formal and informal association in East Timor.

Most Kosovo evaluation reports note the reliance on host families and remittances from abroad, and some (e.g., ICRC/IFRC, 2000) provide quite detailed information on feelings of mutual solidarity, as well as conflict, between host families and refugees. What the affected population was itself able to organise in the Kosovo case clearly outweighed the effects of humanitarian action, a point also made by the DEC with reference to natural disasters in Orissa and Central America (DEC, 2000, 2000b). In the Kosovo case, agencies

failed to support adequately these coping strategies, partly because of their focus on refugee camps (see [Chapter 3](#)). A similar point can be made about the non-Kosovo interventions. A challenge for the future is to understand the different coping strategies of members of the affected population, and build on these during interventions.

#### 4.2.8 Knowledge of context

The evaluation reports demonstrate a major gap in agencies' understanding of the context in which they were intervening. In Kosovo this was one of the most consistent messages from the reports, where the nature of 'ethnic' conflict in particular was inadequately considered. Without an adequate appreciation of context it is difficult, if not impossible, to design appropriate programmes. This is linked in turn to agencies' inadequate attention to coping strategies, where understanding how local and affected populations act in times of crisis is a major factor in understanding local context.

Why do agencies have such difficulty in gaining an understanding of the context in which they are operating? Among likely contributing factors is the lack of analytical tools for use by agencies, though the gap in relation to war economies is now being filled (Le Billon, 2000). Another factor is inadequate use of the capacities of existing local staff. However, the principal factor must be that agencies are simply not according sufficient priority to the development of the necessary analytical capacity, either within their own organisations or elsewhere in the system. Developing the capacity requires an investment that many agencies find difficult to justify or sustain when in direct competition with provision of assistance, or investments in what are perceived as more useful 'sharp end' activities such as logistics.

The reluctance of agencies working in challenging and dangerous environments to make such investments, may highlight the difficulties faced by a system composed of so many separate agencies. Some investments are seen to be uneconomical at individual agency level. The necessary economies of scale may be achieved by a number of agencies sharing costs. One example of such an initiative is the 'Learning Support Office' model (see [Endnote 5, Chapter 3](#)) being developed by ALNAP to provide services to a range of agencies – UN, Red Cross, NGO or others.

#### 4.2.9 Lack of preparedness and rapid response capacity

Another consistent message from the reports is a lack of preparedness and capacity to respond expeditiously to sudden-onset humanitarian needs. In terms of preparedness, this was not just an issue of a lack of information but also of a failure to act on available information. Preparedness does not appear to be an area bilateral donor organisations are overly willing to fund, preferring instead *ex post* interventions. This is perhaps due to the low profile nature of preparedness work and the possibility that funding preparedness may not have results.

Following their experience in Kosovo and elsewhere, several agencies are in the process of setting up rapid response capacities with dedicated personnel or rosters of personnel deployable at short notice, linked in some cases to stockpiled materials. According to the

ECHO evaluation (ECHO, 2000o) UNHCR is currently undertaking a programme which includes the following:

- enhanced systems of personnel recruitment and emergency deployment with 30 experienced personnel members, drawn from HQ and the field, on standby for emergency deployment within 24 to 72 hours;
- set up of an emergency pool of senior managers, to lead or coordinate large-scale and complex emergencies;
- intensive emergency training, especially at the senior and middle-management levels;
- external stand-by arrangements with a number of key partners strengthening emergency reserves;
- a database to determine the emergency capacity of NGOs.

However, agencies have made such investment in the past and, over time, the capacities have been eroded in periods of budget cuts, reorganisations or simply as a result of changed priorities. It remains to be seen whether the agencies will be able to sustain such investments this time around.

Once again the situation presents issues for evaluations. Standards as to what constituted an adequate level of preparedness in a given situation do not seem to have been available to evaluators, with the exception of fairly limited areas, for example, the need to have emergency teams on site within a certain period. The collection and dissemination of good practice case studies would be of assistance, and could be developed into guidance on how to evaluate preparedness.

#### 4.2.10 Inadequate monitoring systems

The attention given by agencies to monitoring is found to be limited in almost all the evaluations. They also consistently lament the lack of systems that could provide data about results. [Chapter 3](#) makes the point that an unmonitored programme is unevaluable in relation to standard performance measures.

Why monitoring is so weak is not clear, but may however relate to the difficulties of establishing effective systems in the first phase of response; the slow trickle down of performance measurement to humanitarian action; the idea that monitoring is not relevant in situations where life is in danger; and a lack of technical skills of agency staff.

In their monitoring recommendations, the reports often repeat evaluation mantras such as ‘set clear objectives’ and ‘develop indicators’, without recognising the constraints to such activities. DRC notes:

‘With events moving so fast all the time in the Kosovo emergency it sometimes appeared pointless to those involved to use indicators of “success” and “failure” since the criteria for them kept changing. What at one stage appeared as a failure might later turn out as an achievement, or in any case as a learning experience,’ (DRC, 1999: p23).

Evaluators and evaluation guidance consistently underestimate the constraints to, for example, setting clear objectives – e.g., the need to negotiate with a range of actors;

political constraints; pressures to allocate and disburse funds; and the need for specific technical skills. Unless constraints are recognised by evaluators and commissioning agencies alike, project managers are unlikely to pay adequate attention to general recommendations relating to monitoring and evaluation systems.

#### 4.2.11 Human resources and management issues

The picture that comes across from most of the reports is of under-trained, under-resourced staff doing their best in difficult conditions. Of all the problems noted, this is one of the most important to rectify. Good performance crucially depends on the ‘right woman or man, in the right place, at the right time’, to paraphrase DFID (2000). The numerous recommendations on human resources fall into three main categories:

- the need for more senior, better qualified personnel to be present at the beginning of the complex emergency/natural disaster;
- the need for personnel to be able to carry out adequate needs assessment including consultation with the affected population, gender and political analysis, and monitoring;
- less staff rotation.

One level of constraints relates to incentive structures within agencies. For example, a general lack of incentives to include gender, or to volunteer for field positions during complex emergencies (UNHCR, 2000a). A second level relates to lack of investment in personnel, including adequate training. In a culture of budget cuts and continuous organisational change, these constraints are difficult to overcome. This is, however, perhaps the area most likely to produce immediate improvements in humanitarian action. Future Annual Reviews might usefully pay attention to capacity development, and report on good practice and advances in this area. They might also monitor whether agencies are meeting their commitments, for example the ECHO Agenda for Action (ECHO, 2000).

People in Aid recently published the audited results of the two-year pilot to implement the People in Aid Code of Best Practice in the *Management and Support of Aid Personnel* (Davidson and Raynard, 2001). Seven of the original twelve agencies were recognised as fulfilling the Code’s conditions. Their commitment to the Code will be re-audited in three years. The framework is, therefore, now established for many more agencies to commit themselves to the Code, and future Annual Reviews will report on the number of agencies making this commitment.<sup>3</sup>

#### 4.2.12 Moral dilemmas in the Kosovo case

Perhaps the major difference between Kosovo and the other complex emergencies/natural disasters was the role of NATO in relief activities and the relationship of humanitarian actors to NATO. This brought greater attention, if not greater focus, to an ongoing dilemma as to the extent to which humanitarian action can be independent from political pressure (ECHO, 2000o, and see discussion in [Chapter 3](#)). Some Kosovo reports consider in detail the moral aspects of cooperation with a military partner (NATO) fighting a war unsanctioned by the UN Security Council (e.g., ECHO, 2000o; MSF, 1999c; DEC, 2000c).

For the most part there is an acceptance that, given geopolitical realities, cooperation was inevitable (e.g., ECHO, 2000o). The focus of many of the reports is thus mainly on the mechanics of cooperation, rather than on ethical and moral issues. ‘Saving lives’ became the justification of cooperation with NATO: ‘The humanitarian community generally accepted that the imperative of saving lives required cooperation with NATO,’ (UNHCR, 2000a: p135). However, even the more sophisticated Kosovo reports, such as the one just quoted, do not appear to have addressed the issue of cooperation in sufficient depth and with sufficient balance.

The unsatisfactory nature of ‘lives saved’ as a non-disaggregated indicator is discussed throughout this *Review*. We need to know, as [Chapter 3](#) points out, whose lives were saved by this cooperation and how? what were the effects of not providing protection to the population inside Kosovo? and, what were the unexpected or unplanned consequences?

A notable exception to the lack of attention to moral and ethical issues is the MSF evaluation that notes while discussing cooperation with NATO:

‘MSF stood out as one of the most outspoken and critical NGOs, but was a lot of times the only one to express concern ... For many NGOs ‘independence’ is not on the agenda. For some it presents a concern but they are trapped on the necessity of institutional financing to be able to survive and are not willing to endanger their position by challenging decisions. Last not least (*sic*), there are NGOs which are perfectly aware of the concept of independence, but give it no importance, what so ever. They have been created as an operational branch of their government,’ (MSF, 1999c: p25).

The point for this Annual Review is not that there is a correct opinion about cooperation with NATO, but, that commissioning agencies should give clearer guidance on the need to pay greater attention to the moral and ethical issues likely to be encountered, when preparing the terms of reference for evaluations.

#### **4.2.13 Shelter and housing**

Housing is noted as one of the main problem areas of intervention in the non-Kosovo reports. In particular they highlight the lack of appropriate material and siting, poor targeting, and poor management and coordination. In contrast, there is surprisingly little critical analysis of the Kosovo housing-related programmes. Information as to the recipients of housing support in the Kosovo intervention, and the use made of that support, is negligible. Targeting issues and the relevance of the five-part categorisation used for allocating housing resources are discussed analytically in Tearfund (2000), but not elsewhere. Otherwise we learn in passing, for example, that many Kosovars own more than one home (ECHO, 2000o), but not the implications of this for the housing programme.

As a consistently problem area across a number of interventions, there is a clear need for a collation of lessons and good practice, as well as for the development of training modules/courses in this area.

#### 4.2.14 Participation

About one-third of the non-Kosovo reports noted a good level of participation of the affected population, at least in the implementation of interventions, and about half of the evaluations considered participation. However, in the Kosovo evaluation reports, little is said about the level of participation. We learn in passing, for instance, that there was little participation of the affected population in the construction of camps by the military (UNHCR, 2000a) and extensive self-help and mutual support in the reconstruction of houses in Kosovo (ECHO, 2000o). Also that (DEC, 2000c: p34): ‘It was not apparent, however, that a priority was given to beneficiary consultation ...’. Again the focus on political and organisational features in the Kosovo evaluations may have excluded attention to the role of the affected population.

Despite the introduction in 1994 of the Red Cross/NGO code of conduct committing signatory agencies to finding ways ‘to involve programme beneficiaries in the management of relief aid’, there remains wide variation in practice and continued criticism on the inadequate involvement of beneficiaries and affected populations. The bulk of humanitarian action is for the most part still a ‘top-down’ affair and ‘downward accountability’ to the beneficiaries and affected populations remains very weak. Recent progress by ALNAP in establishing a project to develop good practice and guidance on consultation with, and participation by, beneficiaries and the affected population in the planning, management, monitoring and evaluation of humanitarian programmes is to be welcomed ([Endnote 1, Chapter 1](#)).

#### 4.2.15 Humanitarian evacuation and transfer programmes

The humanitarian evacuation and transfer programmes (HEP) in the Kosovo conflict were distinct new features. It is hard, however, to judge how relevant these programmes are to the development of policy and practice in the humanitarian system, for as the UNHCR evaluation report notes: ‘The constellation of strategic and political interests that made evacuation programmes possible in this case is unlikely to recur frequently,’ (UNHCR, 2000a: p xii). The same evaluation notes that during the implementation of HEP, UNHCR’s attempts to abide by international agreements and ensure refugees’ rights to appropriate asylum in countries other than that of first asylum, were overruled by bilateral actors for political and visibility reasons.

### 4.3 The quality of Evaluation Reports

#### 4.3.1 Overview

This section is based on the assessment of the 49 individual evaluations against the preliminary ‘quality proforma’ developed by ALNAP for this Annual Review (see [Chapter 1](#) and [Annex 1](#)).

About 20 per cent of reports were considered to be using good practice in most areas covered by the proforma. In the remaining 80 per cent there were many elements of good practice, but on balance the reports were weak in about half the proforma areas.<sup>4</sup> Both

Kosovo and non-Kosovo reports were weaker on the substantive areas of methodological rigour; attention to gender and international standards; and consultation with affected populations. This points to system-wide gaps that need to be addressed by commissioning agencies. Reports were stronger in the provision of contextual information (particularly the non-Kosovo evaluations); relevance and clarity of **terms** of reference; legibility; ease of access; and in crafting recommendations.

Almost all evaluations took a conventional evaluation approach, attempting to combine interviews with agency staff at HQ and in the field; with affected populations (in a minority of cases); document review; and observation of projects. There was little explanation or justification of methodological approaches taken. While there may have been attempts to cross-reference between the different information sources in a few cases, the outcomes were never explained adequately, for example, in terms of how conclusions were reached.

Reports assumed that their approach was ‘objective’, but did not provide discussion of potential bias. Biases were evident across the 49 reports – e.g., lack of attention to gender and affected population consultation. Although no evaluation or assessment, including the present one, can be free from evaluator bias, the point is that biases should be made clear and, in particular, how they have influenced the approach and the analysis of the evaluation.<sup>5</sup>

Another key question raised by the reports, and discussed throughout this Annual Review, is the appropriateness of impact indicators for humanitarian action. The answer seems to be, particularly in the case of complex emergencies, that indicators need to cover both ‘traditional’ areas such as the impact of shelter, food aid and support to health, as well as ‘new’ areas including protection, humanitarian space, human rights and advocacy. In addition, decisions will need to be made as to how to weight these indicators (to avoid the ‘well fed but dead’ syndrome noted in one DANIDA report) to determine overall impact. This is further discussed in the [Next steps](#) section below.

Using Borton and Macrae’s (1997) synthesis as a very approximate baseline, there does appear to have been some progress made in evaluation quality over the last five years, including greater consistency in approach and coverage of key evaluation topics. The OECD-DAC (1999) Guidance, its earlier companion aimed more at field workers (Hallam, 1998) and papers produced by ALNAP itself (liberally referenced and used in the reports) have played a significant role in this improvement. Still, this progress has been from a low base, and much remains to be done (see [4.5](#) below).

The sections below follow the outline of the proforma. More detailed information on the non-Kosovo reports can be found in [Chapter 2](#).

#### 4.3.2 Purpose and focus of evaluation

The majority of the 49 reports have a mixed lesson-learning and accountability focus, usually roughly equal. This duality attempts to meet the needs of different audiences and

terms of reference requirements. However, few of the reports consider the tensions, creative or otherwise, between the two (see [Chapter 1](#)).

### 4.3.3 Constraints

The Kosovo reports note similar constraints to the non-Kosovo set – e.g., lack of data and the complexity of the intervention in relation to the time available to evaluate it. The ECHO Kosovo evaluations and DEC (2000c) note that the refugee emergency was long over by the time the evaluation teams arrived, raising questions of recall and attribution. The reports did not tend to linger on constraints, usually noted in passing. These passing comments, and the few reports that did deal with this issue in more detail, highlight the fact that evaluations of interventions, in response to complex emergencies and natural disasters, face considerable constraints. Some are specific to humanitarian action, such as security issues and lack of access to areas of conflict or disaster. Others are more general in the evaluation and research fields: conflicts among the team or with the commissioning agency; inadequate direction and support from the commissioning agency; reliance on interpreters and agency personnel or the commissioning agency for logistics; and lack of time.

For purposes of transparency and credibility, reports would do well to expand their discussion of these constraints and illustrate their impact on the coverage and outcome of evaluations. For example, Kosovo evaluators must have been almost entirely dependent on interpreters in their discussions with the affected population, but none of the reports makes mention of this as a potential constraint.

### 4.3.4 Terms of reference, team composition, time allowed

Overall terms of reference were adequate where information provided on the nature and objectives of the intervention and clarity of purpose were concerned. Whether team composition was adequate is impossible to say, given the lack of details provided on this. Commissioning agencies tended to draw on a fairly narrow range of institutions and individuals, presumably known quantities. In total, 36 of the individual evaluations were carried out by expatriates, 12 by a mix of host-country citizens and expatriates, and 1 by host country citizens. Of the 55 evaluators involved in the Kosovo evaluations, 52 were expatriates, and 53 were independent consultants (with 2 being agency staff, one from UNICEF and one from WFP).<sup>6</sup> On the basis of the information provided in the reports, it is not possible to say if these consultants had any prior experience in the region, but it can be assumed in the majority of cases that they did not. The lack of contextual understanding and social analysis, for the Kosovo evaluations in particular (see [Chapter 3](#)), is a direct result of this hiring pattern by commissioning agencies. The parallel section in [Chapter 2](#) expands on the discussion of team selection, recommending that a mixed team of expatriates and host-country evaluators should be used.

Allocation of time to undertake the evaluation effectively does not appear to have been adequate in the majority of cases. This appeared to be a particular constraint in relation to consultation with affected populations, undermining overall credibility. Commissioning

agencies should ensure that sufficient time is allowed for consultation with the affected population, and ensure that this consultation takes place effectively. The evaluation team or team leader can be actively involved in planning this element of the terms of reference.

#### 4.3.5 Information on context and intervention

While information on context and the intervention was adequate overall, the Kosovo and non-Kosovo reports display some differences. The Kosovo reports, with some notable exceptions (e.g., DEC, 2000; UNHCR, 2000), are weaker in the provision of context, and, as [Chapter 3](#) notes, presented ‘crises’ without ‘conflict’, and ‘ethnic’ conflicts without adequate analysis of ethnicity. The difference may be accounted for by the fact that the non-Kosovo evaluators were more likely to be specialists in the country/region of intervention.

The reports did not tend to make good use of contextual information. They provided a narrative history of events and actors, but often failed to make necessary linkages between these events and the intervention to illustrate constraints and opportunities. Context was included, perhaps because it is conventional to introduce evaluation reports with it or because it was demanded by the terms of reference. Terms of reference should therefore make clearer the purpose of providing contextual information and the need to link it throughout the analysis. The conclusions of [Chapter 2](#) and [Chapter 3](#), and the assessment against the proforma, also demonstrate the need for evaluation standards in this area, as standards currently differ widely between agencies and individuals.<sup>7</sup>

#### 4.3.6 Methodology and transparency

Few of the reports assessed provided adequate detail on, and justification for, the use of the methodology selected. Given that the choice, development and defence of methodology is a major component of evaluations, this is a significant shortcoming. This is not to say that the methodologies chosen necessarily lacked rigour, only that the reader is not given sufficient information to decide about the appropriateness and strength of methodology. Greater transparency in this area – for example, providing the numbers of affected population consulted (including their gender, socioeconomic grouping and ethnic background) and details of how they were consulted – will add to the evaluation’s credibility (see [4.5](#) below).

Cases of evaluations being potentially biased because they relied mainly on one information source, were rare. Conventional approaches were generally used – e.g., interviews with agency/government staff; document review (although often not integrating the findings from secondary documents into the text); interviews with the affected population; and project visits. There was very little experimentation or use of techniques such as Participatory Rural Appraisal. As noted, while evaluations did use a range of sources, they rarely cross-referenced (triangulated) between them to increase reliability. Evaluation of humanitarian action has some way to go to catch up with techniques in these areas, already commonly used in the evaluation of development cooperation.

Given problems of attribution, it might have been expected that evaluations would make use of control groups, and consultation with the affected population not covered by the intervention, as a means of verifying impact. Only one of the 49 individual evaluations did. The Kosovo conflict would, in particular, appear to be a case where control groups were needed, even if this was made difficult by the reverse movement of refugees back to Kosovo. This is because the affected population did so much for themselves, which means that the impact of external interventions may have been quite limited. Unless we know how well non-recipients of humanitarian action fared, the results of this massive humanitarian programme cannot be plausibly ascertained.

#### 4.3.7 Consultation with beneficiaries and affected population

Beneficiary perspectives are sorely lacking in both sets of reports. Comparatively, there is less heard from the affected population in Kosovo, where only two reports were assessed as being satisfactory or better in this area. This may be because much of the affected population had already returned to Kosovo, while the evaluations focused mainly on operations outside Kosovo. Those evaluations that did focus on operations in Kosovo do not reveal better performance in terms of consultation with the affected population.

Only six evaluation teams, including those undertaking three of the DEC evaluations (DEC, 2000, 2000b, 2000c) consulted adequately with the affected population *and* included adequate methodological details about this consultation. Given the limited information provided in many evaluations, it is impossible to say who was actually consulted; how they were consulted; why they were selected; and what use was made of the information gathered.<sup>8</sup> These evaluations may have consulted adequately, but the reader is often left with statements such as: ‘In the field the team spent as much time as possible with the beneficiaries ...’ (ECHO, 2000), and little else. Even in those cases where there was a good level of consultation, the information gathered was not always integrated into the analysis.

The *idea* of beneficiary consultation is widely accepted and included in most terms of reference. It has however proved difficult to implement. Evaluation teams either fail to allocate enough time and resources or don’t have relevant skills, or while acknowledging its importance find ways of evasion (‘time did not permit’ etc). Consultation with the affected population is one area where commissioning agencies should play a greater gatekeeping role, as well as ensuring that sufficient time has been allowed.

#### 4.3.8 International standards

The conclusion from the evaluation reports and [Chapter 3](#), in relation to international standards, is that their use remains somewhat limited, both in operations and in evaluations. While one third of the Kosovo reports and just under half of the non-Kosovo reports make reference to international standards, these are rarely integrated as evaluation tools to assess performance. [Chapter 3](#) also notes the controversial nature of the standards, and the difficulties of establishing standards across regions and cultures. Evaluations need to make

a distinction between assessing whether agencies have used international standards adequately in their work, and the use of standards to assess the intervention itself.

Evaluations that do integrate standards illustrate their use as evaluation tools and raise important evaluation questions (e.g., DANIDA, 1999c; DEC, 2000b, 2000c; UNHCR, 2000a). (UNHCR, 2000a) provides a field-level example through its use of international conventions and human rights law as a means of assessing UNHCR's performance in protection in the Kosovo conflict, and another is provided by the DEC's Orissa evaluation (DEC, 2000b). The latter used the Sphere Charter to open a dialogue with one NGO, and to evaluate its planned resource allocation to a marginalised group, contrary to the Sphere guidelines.

#### **4.3.9 Attention to gender, the vulnerable or marginalised**

Only about one third of the non-Kosovo evaluation reports contained findings on gender or could be considered to be partly gender-mainstreamed ([Chapter 2](#)). In the Kosovo reports gender is largely absent and one of the biggest gaps, remarkable in that a majority of adult refugees leaving Kosovo appear to have been women (information only imparted in passing). WFP notes that: 'Reports indicated the relocation of many women, children and elderly people from Pristina and southern towns in Kosovo to FYRoM to escape the fighting,' (WFP, 2000b: p33). The absence of attention to gender is even more remarkable given that, in determining that the intervention was both a success and relevant to the needs of the affected population, almost nothing is said about its differential impact on women and men. Part of the reason for the more extensive gaps in the Kosovo evaluations may be that they tended to focus on political issues to the exclusion of cultural and social ones (see [Chapter 3](#)). The lack of adequate attention to gender is one of the most consistent gaps across both sets of evaluations.

In terms of attention to vulnerable or marginalised groups, the non-Kosovo set also fares much better than the Kosovo set with just over half of them covering this area in a satisfactory manner. In the Kosovo set only three paid sufficient attention to these groups. Many of the evaluations focused on geopolitics, internal politics and organisational issues, to the exclusion of a disaggregated analysis of the affected population. Many also display a top-down non-reflective approach in this area that is unlikely to support lesson learning. The views, perceptions and opinions of vulnerable and marginalised groups appear to have been largely ignored in a majority of evaluations, despite a moral obligation on the part of evaluators to represent the interests of this group, and the potential key information they can provide.

#### **4.3.10 Coverage of factors potentially influencing performance**

Reports were assessed as performing reasonably in terms of inclusion of coverage of factors such as geopolitics and the role of local authorities, although significant gaps remain, as pointed out in [4.3.5](#). Because many of the reports, and in particular the Kosovo set, had a focus on institutional performance, this is one of their strongest points. Many evaluators clearly felt more comfortable dealing with management issues such as hiring

practices or head office/field communications, than they did trying to measure impact. This is perhaps not surprising because until recently, and in the last five years in particular, much of the evaluation of development cooperation has focused on intra-institutional issues. It has only been with the introduction of results-based management approaches in many agencies (see [Chapter 1](#)) that there has been an increased focus on measuring impact. Evaluations of humanitarian action appear to have some catching up to do in this area.

#### 4.3.11 **Conclusions and recommendations**

Promoting dialogue about, and sharing the findings of, the draft evaluation has become common practice in the evaluation system. Of the 49 individual evaluations, 37 had adequately shared the draft at least with the commissioning agency, and a number of reports noted how they had responded to comments. This is a positive sign as it increases the likelihood that evaluation report recommendations will be acted on.

Both Kosovo and non-Kosovo reports were relatively strong in terms of the development of clear conclusions, and recommendations that were logically tied to conclusions, however, some were better than others. This is another area where the collection and dissemination of good practice material may be useful, for example where recommendations have been taken up by agencies. ALNAP's work on the follow-up to evaluations (see [Endnote 1, Chapter 1](#)) should produce good practice examples in this area.

#### 4.3.12 **Legibility**

Evaluators have recognised the importance of presenting evaluation findings in an attractive and easy to read format, and this has for the most part been done without sacrificing evaluation quality. Some, however, were less accessible and contained page after page of single-spaced text in small fonts, running in some instances to more than 150 pages. For the most part diagrams and visual aids were well used. Most reports contained an executive summary that accurately and concisely represented the findings of the evaluation. Legibility may thus be seen as a comparative strength of this overall set of evaluation reports.

### 4.4 **Implications of the LRRD Debate for Establishing Evaluation Standards**

Given its particularly problematic nature in relation to the development of evaluation standards and criteria, the LRRD issue is raised again. As USAID notes: 'That relief experts have different views of the purpose of emergency assistance – whether it is for relief only, or for rehabilitation and economic development as well – exacerbates an already complex [methodological] situation,' (USAID, 2000a: p v). How an emergency or disaster is perceived and classified will have a major influence on which indicators of impact are used. While it is common in general evaluation practice to assess interventions against both stated intentions and a wider set of criteria such as impact and sustainability, some evaluation reports argue that these wider criteria are less relevant for the evaluation of humanitarian action (USAID, 2000; ECHO, 2000).

However, the OECD-DAC (1999) guidance for evaluation of complex emergencies includes the criteria of impact and sustainability/connectedness (see [Endnote 3, Chapter 2](#), among the areas to be evaluated. It thus requires an evaluation against broader criteria than just intervention objectives, which it defines as effectiveness. In the general evaluation field, impact is usually considered a ‘higher-level’ indicator of results, and it may be that in the determining of the impact of humanitarian action, short-term and long-term objectives should be equally weighed. This would indicate a shift away from current practice, where short-term objectives (such as saving lives) are used as primary indicators. By asking fundamental questions about the results of an intervention, evaluation feedback mechanisms should help establish priorities and stimulate debate in this area.

The fact that many of the evaluations cover humanitarian action and rehabilitation, and in some cases development, has implications for evaluators and evaluation offices. Support documents (handbooks, good practice guidelines, etc) need to focus not only on crisis situations and relief, but also on how to evaluate the link to development.

#### **4.5 Next Steps in Evaluation of Humanitarian Action**

While there is much good practice in both Kosovo and non-Kosovo evaluation reports, and there appears to have been progress in some areas in terms of quality, much remains to be done. There is also a pressing need to learn from the wider evaluation field since the evaluations considered in this Annual Review have hardly drawn on lessons from this source. The quality of evaluations currently accepted by commissioning agencies is too low in several areas. There is a pressing need for the development of evaluation of humanitarian action standards, tools and capacity. These will help improve evaluation quality and credibility, as well as providing a means of holding evaluators accountable.

Development of standards and tools should adopt a dual focus, first in relation to the topic areas of impact, relevance, connectedness, etc, and second in relation to the ‘new’ agenda items such as protection and advocacy. Evaluators need clearer guidelines concerning what is expected from them in both areas, as well as the linkages between them. The relative weight to be assigned to each in order to determine the impact of the intervention also needs to be clarified. The majority of evaluations, whose purpose was to account for expenditure of material inputs, were perceived as apolitical. Yet, widening their perspectives – i.e., factoring in aspects such as protection and human rights, both more politically sensitive – has system-wide implications in need of discussion. Can impartiality of evaluations be assured when assessing politically sensitive factors? The evaluation of protection also raises the issue of what expertise and experience there is in the analysis of protection in the evaluation community and, more particularly, in the case of evaluators from countries and regions primarily affected by complex emergencies and natural disasters.

The development of evaluation standards could follow the process that led up to the production of the US Program Evaluation Standards (Joint Committee, 1994), which, to the

authors' knowledge, are currently the main set of standards in use in the wider evaluation field.

'The Joint Committee developed a systematic public process for establishing and testing the new standards before recommending their use ... the process involved many experts in evaluation, users of evaluation, and others ... Several hundred educators, social scientists, and lay citizens were involved in the development of the first program evaluation standards,' (ibid. p xvii).

Increased participation is likely to provide a wider range of experience and ensure buy-in. It is anticipated that the development of standards for the evaluation of humanitarian action might be a two- to three-year process at the least.

'Standard A4', an example which illustrates the approach of the 30 Program Standards, covers 'Defensible Information Sources' as follows: 'The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed,' (ibid, p141). Clearly, the quote under [Consultation with beneficiaries](#) ('In the field the team spent as much time as possible with the beneficiaries,') would not meet the standard. The Joint Committee (1994) also provides an overview of the standards, guidelines, common errors and illustrative cases.

Specific tools such as the proforma used in this Review can be developed in parallel, again in a participatory fashion, and could include tools for the assessment of protection. Good practice material could also be disseminated on, for example, the provision of contextual information and the crafting of recommendations.

In the meantime immediate steps could be taken, with minimal effort and cost, to increase the transparency and credibility of evaluations, all related to the provision of more information:

- the commissioning agency can note in an introduction how the evaluation team was selected, and the intended circulation and use of the evaluation;
- the evaluation can note constraints and show how these have impacted on data gathering and analysis;
- the evaluation can provide details of the methodology, and in particular disaggregated information on consultation with the affected population, including numbers consulted, and methods used for consultation and analysis of data;
- the evaluation can provide a short biography of team members, noting their regional and sectoral expertise, any work in the gender field, their language capacity and their experience in evaluation (e.g., DEC, 2000);
- all data should be gender-disaggregated where possible.

Finally, there were instances where the same agency had almost simultaneously commissioned evaluations where one had been strong and the other weak in exactly the same area. This suggests commissioning agencies can improve on their role as gatekeepers of evaluation standards and practice, particularly in relation to identified gaps.

## 4.6 Concluding Remarks

This Annual Review has demonstrated the inherent value of bringing together and systematically synthesising the findings of a large set of evaluations of humanitarian action. By highlighting that these are system-wide performance issues, whether areas found wanting or good practice deserving emulation, it has confirmed their relevance to all within the international humanitarian system. Its principal findings should, therefore, help set the agenda for ongoing discussions on how to improve performance.

The quality assessment of the evaluation reports, through meta-evaluation, also reveals a substantial agenda for those involved in the commissioning, managing, undertaking and use of evaluations of humanitarian action.

The Review makes a number of criticisms of current evaluation practice and standards in some of the more important and substantive areas, including methodological rigour and transparency. This is done in the context of the need to improve evaluation practice, as evaluation and results-based management are here to stay for the medium-term future at least (if only due to the requirement that agencies illustrate adequate results to funders and governments).

The precise form of evaluation of humanitarian action is still under development, and some have questioned whether the evaluation mechanism needs to be complemented by social audits (Raynard, 2000). It is clear, however, that if sufficiently rigorous and adherent to good practice standards, evaluation of humanitarian action has the potential to make a significant contribution both to accountability and to lesson learning, leading to improved practice.

## ENDNOTES

### Chapter 1

1. As well as the Annual Review series, ALNAP's 2001–02 Workplan activities in relation to the evaluation of humanitarian action include: the publication of a volume of reflections by evaluation practitioners on nine humanitarian action evaluations undertaken between 1993 and 2000 (Wood et al, 2001); the development of short courses and training modules for agency personnel and evaluators; studies on how evaluations of humanitarian action are used and followed up by agencies; and the sharing of evaluation plans between members to reduce duplication and encourage greater collaboration in evaluation. Other accountability and learning activities include a book on accountability frameworks in the international humanitarian system; a major international collaborative study to develop guidance for humanitarian agencies on consultation with and participation of beneficiaries and the affected population; and testing the field level 'Learning Support Office' (see [Endnote 5, Chapter 3](#)).
2. The ALNAP symposium 'Learning from Evaluation: Humanitarian Assistance and Protection in Kosovo' was held in Geneva on 17 & 18 October 2000. Some 51 representatives (military/Red Cross movement /NGO/bilateral and multilateral donor organisations/UN plus academics and consultants) from 43 different organisations that had been involved in the Kosovo crisis or the evaluation of humanitarian programmes, participated in the event.
3. Operation Provide Comfort was mounted in the aftermath of the Gulf War between Coalition forces and the Iraqi forces that had occupied Kuwait. Following the March 1991 ceasefire between the Coalition forces and the Iraqi government, a popular uprising took place in the predominantly Kurdish areas of northern and eastern Iraq. The Iraqi military's brutal suppression of the uprising led to the exodus of almost 2 million Kurds towards Iran and Turkey. The Turkish authorities refused them admission and almost 500,000 Kurds were trapped in the mountains along the Iraqi/Turkish border. Following Security Council Resolution 688 of 5 April 1991, US and European airforces began airdropping supplies and US, British and French forces created a 'safe haven' in northern Iraq, enabling the provision of direct humanitarian assistance by the militaries, the UN, the Red Cross and NGOs and the movement of the Kurds down from the mountains back into Iraq. The operation was a watershed event for the humanitarian system as it set a precedent for military intervention in support of humanitarian objectives on sovereign territory that was subsequently repeated in Somalia, Kosovo and East Timor.
4. ICRC, IFRC, UNICEF, UNHCR and WFP are the five organisations involved in the benchmarking study.
5. The DAC Working Party on Aid Evaluation maintains a database of evaluations of development cooperation undertaken by bilateral and multilateral donor organisations <[www.oecd.org/dac/evaluation](http://www.oecd.org/dac/evaluation)>. Whilst the database is extremely valuable, evaluations undertaken by organisations other than DAC members, in the international humanitarian sector are not included and the number of evaluations of humanitarian action is limited.
6. See [Chapter 2 Endnote 4](#).
7. The book is scheduled for publication by Zed Press in August 2001.
8. The proforma has been given a blind pre-test by two independent consultants to assess consistency of application across different assessors and help guide the proforma's further development/refinement. The objective is to strengthen its methodological basis to ensure its acceptance.

## Chapter 2

1. Use has also been made of a paper (HPN, 2000) that compares the approaches and results of an evaluation included in the set (DEC, 2000) with two evaluative studies of humanitarian operations in the same region, not included in the set (Groupe URD, 2000; Lister, 2000).
2. The DANIDA evaluation reports cover the period 1992–99, otherwise most reports cover (1999 and 2000).
3. Connectedness concerns the need for humanitarian programmes to take longer term needs into account in addition to their focus on addressing immediate, acute needs. The criteria is similar to the standard evaluation criteria of ‘sustainability’ but takes account of the reality that the majority of humanitarian programmes are not intended to be sustainable (OECD-DAC, 1999).
4. Triangulation is a key technique allowing cross-referencing and cross-checking of different information sources. In complex emergencies/disasters where data may in general be poor, triangulation will support evaluation rigour.
5. Part of UNHCR’s response, to the ECHO (2000b) evaluation of support to rehabilitation programmes in Rwanda, was to note that the evaluation had paid insufficient attention to the protection aspects of the intervention.
6. HPN and ALNAP literature was widely used in the evaluations, both for the development of methodology and for comparative purposes.
7. The question of ‘impartiality’ is under discussion in the general evaluation field at present, where there has been a questioning of the concept of the evaluator as an objective and unbiased observer. Related to this is the concept that evaluations, to be effective and used, need to be as participatory as possible.
8. The latter two figures are estimates based on the schedules in the reports.
9. A control group is a group chosen randomly from the same population as the group that benefited from the programme but for some reason the control group did not benefit from the programme. A control group therefore enables an assessment to be made of what would probably have happened to the programme group if the programme had not been undertaken.
10. In one report the area was not applicable.
11. Interview with Ed Tsui, Director, Policy, Advocacy and Information Division, UNOCHA, February 2001.

## Chapter 3

1. A much shorter, but structurally similar, form of this paper was presented and discussed at the ALNAP symposium ‘Learning from Evaluation: Humanitarian Assistance and Protection in Kosovo’, which took place in Geneva from 17–18 October 2000. The title of that paper was ‘Kosovo Humanitarian Programme Evaluations: Towards Synthesis, Meta-evaluations and Sixteen Propositions for Discussion’. It is now presented, in greater depth as Chapter 3. This takes into account discussions from the plenary sessions which focussed on building immediate response capacity, improving civil military cooperation, achieving social learning, and reconceptualising humanitarian programme evaluation as critical learning practice. Evaluations not available earlier were obtained and considered.  
It is important to emphasise that the purpose of this paper is to provide an overview of common (and dissenting) themes and messages, not to highlight one evaluation or study over another.
2. See Chapter 1 Endnote 2.
3. *See* (Scriven, 1991).
4. *See* (Pottier, 1999).

5. The objective of the ALNAP Learning Support Office project is to make a positive impact on the quality of emergency response in the field through the promotion and facilitation of three-way learning activities: i. 'learning-in'; ii. 'learning-out'; and iii. 'lateral learning'. It is proposed that a test Learning Support Office will be run in Sierra Leone in 2001.

## Chapter 4

1. Work to develop criteria for the evaluation of protection activities which draws on those so far developed by UNHCR and ICRC is planned by ALNAP as part of its Evaluation 'Guidance Gap Filling' activities.
2. Information on the Network is available on the Global Peacebuilding Network site at <http://wbln0018.worldbank.org/ESSD/pc1.nsf/Home>
3. Information on People in Aid is available at [www.peopleinaid.org](http://www.peopleinaid.org)
4. These figures are necessarily approximate given the aggregate nature of the assessments.
5. In the current Review there are clear biases, e.g., in terms of attention to protection, and in the choice of topics for the proforma. Further biases are spelt out in [Chapter 3](#).
6. Some evaluations (e.g., ICRC/IFRC, 2000; ECHO, 2000p) used local research teams from the agency being assessed, but the involvement of these teams, or the implications of this for the evaluation, is not noted in detail.
7. The Joint Committee on Standards for Educational Evaluation (Joint Committee, 1994: p133) gives as the standard for 'Context Analysis': 'The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.'
8. This is in contrast to the long lists of agency staff consulted that usually made up an Annex to the evaluations.

## ANNEX 2

### ACRONYMS

<b>ACT</b>	Action by Churches Together
<b>ALNAP</b>	Active Learning Network for Accountability and Performance in Humanitarian Action
<b>AUSAID</b>	Australian Agency for International Development
<b>BADC</b>	Belgian Administration Development Cooperation
<b>BRCS</b>	British Red Cross Society
<b>CIDA</b>	Canadian International Development Agency
<b>CIMIC</b>	Civil Military Co-operation
<b>DAC</b>	Development Assistance Committee
<b>DANIDA</b>	Danish International Development Assistance
<b>DART</b>	Disaster Assistance Response Team
<b>DEC</b>	Disasters Emergency Committee
<b>DFID</b>	Department for International Development (UK)
<b>DMI</b>	Disaster Mitigation Institute
<b>DRC</b>	Danish Refugee Council
<b>EAPC</b>	Euro-Atlantic Partnership Council
<b>ECHO</b>	European Community Humanitarian Office
<b>EMG</b>	Emergency Management Group (Albania)
<b>EMOP</b>	Emergency Operation
<b>ERD</b>	ALNAP Evaluative Reports Database
<b>EU</b>	European Union
<b>FAO</b>	Food and Agriculture Organisation (UN)
<b>FYRoM</b>	Former Yugoslavia Republic of Macedonia
<b>GAA</b>	German Agro Action
<b>Groupe URD</b>	Groupe Urgence-Réhabilitation-Développement
<b>GTZ</b>	Gesellschaft für Technische Zusammenarbeit
<b>HAP</b>	Humanitarian Accountability Project
<b>HEP</b>	Humanitarian Evacuation Programme
<b>HPN</b>	Humanitarian Practice Network (ODI)
<b>HRM</b>	Human Resource Management
<b>HTP</b>	Humanitarian Transfer Programme
<b>IASC</b>	Inter-Agency Standing Committee
<b>ICG</b>	International Crisis Group
<b>ICRC</b>	International Committee of the Red Cross
<b>ICVA</b>	International Council of Voluntary Associations
<b>IDP</b>	Internally Displaced Person

<b>IFRC</b>	International Federation of Red Cross and Red Crescent Societies
<b>INGO</b>	International Non-governmental Organisation
<b>JEEAR</b>	Joint Evaluation of Emergency Assistance to Rwanda
<b>JICA</b>	Japan International Cooperation Agency
<b>KLA</b>	Kosovo Liberation Army
<b>LRRD</b>	Linking Relief, Rehabilitation and Development
<b>M &amp; E</b>	Monitoring and Evaluation
<b>MSF</b>	Medecins sans Frontieres
<b>MSF-B</b>	Medecins sans Frontieres - Belgium
<b>MSF-H</b>	Medecins sans Frontieres - Holland
<b>NATO</b>	North Atlantic Treaty Organisation
<b>NGO</b>	Non-governmental Organisation
<b>ODI</b>	Overseas Development Institute
<b>O &amp; M</b>	Operations and Management
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>OFDA</b>	Office of Foreign Disaster Assistance (USAID)
<b>OSCE</b>	Organisation for Security and Co-operation in Europe
<b>PIA</b>	People In Aid
<b>SCHR</b>	Steering Committee for Humanitarian Response
<b>SCF</b>	Save the Children Fund
<b>SDC</b>	Swiss Development Cooperation
<b>SIDA</b>	Swedish International Development Agency
<b>TOR</b>	Terms of reference
<b>UNDP</b>	United Nations Development Programme
<b>UNHCHR</b>	United Nations High Commissioner for Human Rights
<b>UNHCR</b>	United Nations High Commission for Refugees
<b>UNICEF</b>	United Nations Children's Fund
<b>UNOCHA</b>	United Nations Office for the Coordination of Humanitarian Assistance
<b>UNTAET</b>	United Nations Transitional Administration in East Timor
<b>USAID</b>	United States Agency for International Development
<b>VOICE</b>	Voluntary Organisations in Cooperation in Emergencies

**WFP**  
**WHO**

World Food Programme (UN)  
World Health Organisation (UN)

## ANNEX 3

### 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.

**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.

**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. See [Chapter 1](#) for characteristics of evaluation of humanitarian action.

**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** ‘Action undertaken for the advancement of the welfare of humanity without regard to race, religion or politics,’ (Gunn, 1990).

**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.

**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 being able to provide assistance on a genuinely non-discriminatory basis.

**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.

**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).

**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-DAC, 2000: p16).

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

# ANNEX 4

## REFERENCES FOR SOURCE EVALUATIONS

### by Commissioning Agency with Key Descriptors

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

**Focus:** Main sectors/issues covered in the evaluation.

**Criteria:** Evaluation criteria used to assess the humanitarian action being evaluated.

**Comments:** Any additional information not covered above.

## Non-Kosovo Evaluations from 1999–2000 (Chapters 2 & 4)

### DANIDA

**DANIDA (1999a) *Evaluation of Danish Humanitarian Assistance 1992–98: Volume 1, Synthesis Report* (Copenhagen: DANIDA) ETC UK**

**Objective:** Synthesise the results of six country case studies and two thematic studies to assess and document Danish humanitarian assistance between 1992 and 1998.

**Focus:** Policy, coordination, funding, strategy.

**Criteria:** Relevance, effectiveness, efficiency, impact, coherence, connectedness.

**DANIDA (1999b) *Evaluation of Danish Humanitarian Assistance 1992–98: Volume 2, Afghanistan* (Copenhagen: DANIDA) Overseas Development Institute (ODI)**

**Objective:** Assess and document the performance of Danish humanitarian assistance to Afghanistan.

**Focus:** Health, mine action, advocacy, UN coordination, NGO coordination, management, protection, security, food aid, refugees, IDPs, rehabilitation.

**Criteria:** Appropriateness, relevance, effectiveness, cost-effectiveness, impact, coverage, efficiency, connectedness, coherence.

**DANIDA (1999c) *Evaluation of Danish Humanitarian Assistance 1992–98: Volume 3, Angola* (Copenhagen: DANIDA) T&B Consult**

**Objective:** Assess and document the performance of Danish humanitarian assistance to Angola.

**Focus:** Rehabilitation, reintegration, mine action, coordination, refugees, IDPs, health, policy, preparedness, protection.

**Criteria:** Impact, coverage, effectiveness, relevance, efficiency, connectedness, coherence, cost-effectiveness.

**DANIDA (1999d) *Evaluation of Danish Humanitarian Assistance 1992–98: Volume 4, Caucasus* (Copenhagen: DANIDA) COWI**

**Objective:** Assess and document the performance of Danish humanitarian assistance to the Caucasus.

**Focus:** Food aid, health, shelter, coordination, logistics, security, preparedness, advocacy, policy, refugees, IDPs.

**Criteria:** Relevance, effectiveness, efficiency, impact, coverage, coherence, sustainability.

**DANIDA (1999e) *Evaluation of Danish Humanitarian Assistance 1992–98: Volume 5, Former Republic of Yugoslavia* (Copenhagen: DANIDA) Landell Mills Management Consultants**

**Objective:** Assess and document the performance of Danish humanitarian assistance to FRY.

**Focus:** Coordination, shelter, refugee camps, rehabilitation, education, psychosocial, policy.

**Criteria:** Relevance, efficiency, effectiveness, impact, coverage, coherence, connectedness,

**DANIDA (1999f) *Evaluation of Danish Humanitarian Assistance 1992–98: Volume 6, Great Lakes* (Copenhagen: DANIDA) CHANNEL Research Ltd.**

**Objective:** Assess and document the performance of Danish humanitarian assistance to the Great Lakes Region.

**Focus:** Refugees, IDPs, coordination, policy, protection, food aid, shelter.

**Criteria:** Relevance, effectiveness, efficiency, impact, connectedness, coherence, coverage.

**DANIDA (1999g) *Evaluation of Danish Humanitarian Assistance 1992–98 : Volume 7, Sudan* (Copenhagen: DANIDA) Overseas Development Institute (ODI)**

**Objective:** Assess and document the performance of Danish humanitarian assistance to the northern and southern sectors of Sudan.

**Focus:** Coordination, management, policy, famine, conflict, preparedness.

**Criteria:** Appropriateness, relevance, coverage, connectedness, impact, efficiency, effectiveness.

**DANIDA (1999h) *Evaluation of Danish Humanitarian Assistance 1992–98 : Volume 8, UN and international organisations* (Copenhagen: DANIDA) Chr. Michelsen Institute**

**Objective:** Evaluate five central policy issues that have faced UN agencies and international humanitarian organisations funded by Danish aid during the 1990s.

**Focus:** System-wide coordination, policy, early warning systems, preparedness, local capacity building, LRRD, protection.

**Criteria:** Relevance, coherence, efficiency, effectiveness, impact.

## **Department For International Development, UK (DFID)**

**DFID (1999) *An Evaluation of HMG's Response to the Montserrat Volcanic Emergency: Volumes I & II* (London: DFID) Barrow, C., Benson, C., Clay, E., Dempster, J., Kokelaar, P., Pillai, N. and J. Seaman**

**Objective:** Review HMG's preparations, organisation and delivery of emergency aid in response to Montserrat crisis.

**Focus:** Shelter, food, education, management, preparedness.

**Criteria:** Impact, appropriateness, cost-effectiveness, coverage, coherence, connectedness, effectiveness.

## **Disasters Emergency Committee (DEC)**

**DEC (1999) *Evaluation of Expenditure of Sudan Appeal Funds* (London: DEC) Corbett, M., Kejji, E., Kirkby, J., Jerome, M., O'Keefe, P. and F. Wekesa**

**Objective:** Analyse expenditure of appeal funds and assess effectiveness, strengths and weaknesses of projects funded.

**Focus:** Nutrition, famine relief, health, watsan, logistics, coordination.

**Criteria:** Impact, effectiveness, efficiency, relevance.

**DEC (2000) *Independent Evaluation of Expenditure of DEC Central America Hurricane Appeal Funds* (London: DEC) Espacios Consultores**

**Objective:** Assess overall appropriateness, effectiveness and impact of a sample of projects supported by DEC funds.

**Focus:** Food aid, shelter, watsan, health, food security, beneficiaries.

**Criteria:** Impact, cost-effectiveness, coverage, appropriateness, effectiveness.

**DEC (2000a) *Bangladesh 1998 Flood Appeal: An Independent Evaluation* (London: DEC) Young, R.**

**Objective:** Assess the appropriateness and effectiveness of the activities financed by the DEC appeal.

**Focus:** Food aid, health, shelter, watsan, coordination, preparedness, beneficiaries.

**Criteria:** Appropriateness, impact, effectiveness, efficiency, coherence, coverage.

**DEC (2000b) *Independent Evaluation of Expenditure of DEC India Cyclone Appeal Funds: Volumes 1 & 2* (London: DEC) Dash, D., Goyder, H., Hunt, M., Missal, L., Padhi, T. and D. Rayanna**

**Objective:** Evaluate the overall impact of the £5 million spent by the 11 DEC members on relief and rehabilitation.

**Focus:** Preparedness, coordination, shelter, agriculture, health, watsan, FFW, livelihoods, advocacy, gender.

**Criteria:** Appropriateness, efficiency, impact, coverage, connectedness, coherence.

## **European Community Humanitarian Office (ECHO)**

**ECHO (1999) *Evaluation of the Humanitarian Aid Funded by the European Commission between 01.01.91 and 30.06.96. Phase III – Final Version (Synthesis)* (Brussels: ECHO)**

**Objective:** Assess EC humanitarian aid, January 1991–June 1996.

**Focus:** Policy, LRRD, coordination, human resources, preparedness, organisational structure.

**Criteria:** Impact, relevance, effectiveness, efficiency, sustainability, complementarity, coherence.

**ECHO (2000) *ECHO Humanitarian Aid: An Analysis of ECHO Evaluations 1996–99. Synthesis Report. ECHO Evaluation Report No 280.* (Brussels: ECHO)**

**Objective:** Synthesise the analysis of the 93 ECHO evaluation reports written between 1996 and 1999.

**Focus:** ECHO aid, coordination, funding.

**Criteria:** Relevance, efficiency, effectiveness, cost-effectiveness, impact and sustainability.

**ECHO (2000a) *ECHO Humanitarian Aid: An Analysis of ECHO Evaluations 1996–99. ECHO Evaluation Report No 279.* (Brussels: ECHO) Eijkenaar, J., Hummelink, M. and J. Telford**

**Objective:** Analyse 93 ECHO evaluation reports carried out between 1996 and 1999.

**Focus:** Findings of 8 separate criteria based reports looking at all aspects of ECHO's operations.

**Criteria:** Relevance, efficiency, effectiveness, cost-effectiveness, impact, sustainability.

**Comments:** Covers recommendations on how to ensure follow-up on evaluation results.

**ECHO (2000b) *ECHO Evaluation Report Global Plans for the Great Lakes Region 1998–99: Rwanda Rehabilitation – UNHCR* (Brussels: ECHO) Eijkenaar, J. and S. de Boer**

**Objective:** Assess ECHO funded UNHCR rehabilitation programme.

**Focus:** Rehabilitation, resettlement, shelter, water, infrastructure, protection, coordination, LRRD, beneficiaries, security.

**Criteria:** Relevance, coherence, effectiveness, efficiency, impact, connectedness.

**ECHO (2000c) *ECHO Evaluation Report Global Plans for the Great Lakes Region 1998–99: Rwanda – Shelter – German Agro Action (Brussels: ECHO)*** Eijkenaar, J. and S. de Boer.

**Objective:** Assess ECHO-funded GAA shelter programme.

**Focus:** Shelter, coordination, logistics, operational capacity, security, water, infrastructure, LRRD.

**Criteria:** Relevance, coherence, effectiveness, cost-effectiveness, efficiency, impact.

**ECHO (2000d) *ECHO Evaluation Report Global Plans for the Great Lakes Region 1998–99: Tanzania Multisectoral Support to Refugees – UNHCR (Brussels: ECHO)*** Eijkenaar, J. and E. Brusset

**Objective:** Assess ECHO funded UNHCR refugee operations.

**Focus:** Refugee camps, preparedness, coordination, security, LRRD, partner capacity.

**Criteria:** Relevance, effectiveness, cost-effectiveness, efficiency, impact, sustainability, coherence.

**ECHO (2000e) *North Mali and North Niger Water and Rural Development: ECHO Evaluation, Global Plans 1996–99 (Brussels: ECHO)*** de Boer, S., Seydoux, L. and R. Sole

**Objective:** Analyse appropriateness and success of ECHO aid to reintegration, sustaining livelihoods and peacebuilding.

**Focus:** Water, agriculture, food security, strategy, LRRD, coordination, security.

**Criteria:** Relevance, coherence, complementarity, effectiveness, efficiency, impact, sustainability.

**ECHO (2000f) *Evaluation of the Humanitarian Aid in Favour of the Cuban Population 1999–2000 (Brussels: ECHO)*** Artundo, C. and M. Coenegrachts

**Objective:** Assess relevance and coherence of ECHO intervention and recommend an exit strategy.

**Focus:** Health, social services, infrastructure, non-food aid, coordination, strategy, gender, LRRD.

**Criteria:** Relevance, coherence, complementarity, effectiveness, efficiency, impact.

**ECHO (2000g) *ECHO Evaluation Report Global Plans for the Great Lakes Region 1998–99: Tanzania Food Sector – WFP (Brussels: ECHO)*** Eijkenaar, J. and E. Brusset

**Objective:** Assess ECHO funded WFP food sector operations.

**Focus:** Food aid, refugees, coordination, LRRD.

**Criteria:** Relevance, coherence, efficiency, cost-effectiveness, effectiveness, impact, sustainability.

**ECHO (2000h) *ECHO Flight Evaluation. Transport in Support of ECHO and DG-DEV Operations. ECHO Evaluation Report 297. (Brussels: ECHO)*** Kunze, M. and D. Tompkins

**Objective:** Provide information to decide the most cost-effective way of providing safe and efficient transport to support ECHO and DG DEV operations in the Horn of Africa.

**Focus:** Transport, logistics, coordination.

**Criteria:** Relevance, effectiveness, cost-effectiveness, efficiency, complementarity, impact.

**ECHO (2000i) *Evaluation Food Sector 1998–99 and 1999–2000 ECHO Tajikistan (Brussels: ECHO)*** Hoogendoorn, A.

**Objective:** Review execution and impact of ECHO-funded food activities in Tajikistan since April 1998.

**Focus:** Nutrition, public health, coordination, food aid, food-for-work.

**Criteria:** Efficiency, cost-effectiveness, effectiveness, coherence, impact, coverage.

## MSF Holland (MSF-H)

**MSF (1999) *The MSF Cholera Intervention in Uganda: December 1997–September 1998* (Amsterdam: MSF-H) Sondorp, E.**

**Objective:** Document the intervention and lessons learned, input to policy, planning and the revision and expansion of MSF Cholera Guidelines.

**Focus:** Health, planning, implementation, epidemics, policy.

**Criteria:** Relevance, appropriateness, effectiveness, impact, cost-effectiveness, efficiency, coverage, coherence, connectedness.

**MSF (1999a) *The MSF Holland Programmes to Support Basic Health Services in Malange and Zaire Provinces, Angola* (Amsterdam: MSF-H) Chabot, J. and G. Kuperus**

**Objective:** Evaluate MSF's policy and performance in supporting district health services.

**Focus:** Health, maternal health, preparedness, human resources.

**Criteria:** Appropriateness, effectiveness, coverage, connectedness, coherence, efficiency.

**MSF (1999b) *The MSF Natural Disaster Preparedness Project in Bangladesh: February 1998–November 1998* (Amsterdam: MSF-H) Van de Putte, B.**

**Objective:** Evaluate MSF-H response to 1998 floods with view to further development of planning for Bangladesh natural disaster preparedness.

**Focus:** Preparedness, floods, policy, health, coordination.

**Criteria:** Appropriateness, effectiveness, connectedness, coherence, coverage, efficiency.

## Norwegian Ministry of Foreign Affairs

**Norwegian MFA (1999) *Evaluation of Norwegian Support to Psycho-Social [sic] Projects in Bosnia-Herzegovina and the Caucasus: Evaluation Report 3.99* (Oslo: Norwegian Ministry of Foreign Affairs) Agger, I., Herzberg, A., Jareg, E., Mimica J. and C. Rebien**

**Objective:** Evaluate experience of psychosocial programmes run by NGOs to develop future policy for support to such programmes.

**Focus:** Psychosocial, policy, capacity building, management, coordination.

**Criteria:** Impact, connectedness, appropriateness, effectiveness, relevance, coherence.

## OXFAM

**Oxfam (2000) *Evaluation of the Integrated Humanitarian Programme in Uraba, Colombia* (London: Oxfam) Botero, A., Prieto, F., Rodriguez, O. and M. Turnbull**

**Objective:** Identify, measure and evaluate the results and impact of the programme.

**Focus:** Food aid, health, psychosocial, advocacy, protection, coordination, management, LRRD.

**Criteria:** Impact, effectiveness, efficiency, relevance, flexibility, connectedness.

## Save the Children Fund, UK (SCF)

**SCF (2000) *Final Evaluation: The Darfur Early Warning and Food Information System* (London: SCF) Bush, J. and J. Seaman**

**Objective:** Review whether project meets its objective of maintaining food security in Darfur area.

**Focus:** EWS, food security, needs assessment, coordination.

**Criteria:** Efficiency, effectiveness, impact, appropriateness, sustainability, coverage.

## United Nations Development Program (UNDP)

**UNDP (2000) *Sharing New Ground in Post-Conflict Situations: The Role of UNDP in Support of Reintegration Programmes* (New York: UNDP)** Barnes, S., Brucchaus, E., Kelly, J., Kulesa, M., Ormsby, G., Rahman, T. and L. Silva

**Objective:** Analyse globally UNDP's involvement in reintegration operations regarding IDPs, refugees and ex-combatants.

**Focus:** Policy, coordination, programming, capacity building.

**Criteria:** Effectiveness, coherence.

## United Nations High Commission for Refugees (UNHCR)

**UNHCR (2000) *External Evaluation of the UNHCR Shelter Program in Rwanda 1994–1999* (Geneva: UNHCR)** Bugnion, C., Gacuruzi, B., Laurent, C. and F. Twagirayezu

**Objective:** Assess the shelter programme's impact on the stability and sustainability of the post-conflict return and reintegration process.

**Focus:** Shelter, reintegration, LRRD, coordination, security, protection.

**Criteria:** Impact, coverage, connectedness, effectiveness, relevance, efficiency, coherence.

## United Nations Transitional Administration in East Timor (UNTAET)

**UNTAET (2000) *East Timor Consolidated Appeal Process (CAP) Review Phase 3 : External Review* (Dili: UNTAET)** Bugnion, C., da Conceicao, C., de Costa, H., Donnelly, J., McDowell, C., Naganawa, S., Niwa, N. and J. Rock

**Objective:** Assess appropriateness and effectiveness of the humanitarian programme.

**Focus:** Coordination, food aid, shelter, health, IDPs, refugees, livelihoods, beneficiary participation.

**Criteria:** Impact, effectiveness, appropriateness, coherence.

**Comments:** The evaluation draws on the supporting beneficiary study and agency self-assessment (UNTAET, 2000a, 2000b).

## US Agency for International Development (USAID)

**USAID (2000) *Hurricane Mitch: Management Assessment of Humanitarian Assistance Activities in Honduras and Nicaragua* (Washington: OFDA/USAID)** Adelski, E., Nurick, K., Rowe, S., Painter, F. and M. Seirig

**Objectives:** Assess quality and effectiveness of USAID's response to affected areas in aftermath of hurricane Mitch.

**Focus:** Preparedness, management, health, nutrition, shelter, watsan, coordination, communication.

**Criteria:** Impact, effectiveness, appropriateness.

**USAID (2000a) *Complex Humanitarian Emergencies and USAID's Humanitarian Response* (Washington: CDIE/USAID) Adelski, E., Hill, R., Mason, J., McClelland, D. and R. Muscat**

**Objective:** Assess effectiveness of U.S. emergency assistance in response to complex humanitarian emergencies.

**Focus:** Policy, coordination, LRRD, food aid.

**Criteria:** Effectiveness, impact.

**Comments:** Based on case studies in Haiti, Mozambique, Rwanda and Bosnia-Herzegovina.

## **World Food Programme (WFP)**

**WFP (1999) *Emergency Food Assistance to Households Affected by Hurricane Mitch : Full Report of the Evaluation of Central America EMOP 6079* (Rome: WFP) Alvarez, E., Kirkby, J., Rubin, F. and A. Sandoval**

**Objective:** Assess impact of regional emergency operation (EMOP) and guide implementation of subsequent protracted relief operation.

**Focus:** Nutrition, health, communication, coordination, security.

**Criteria:** Effectiveness, efficiency, coverage, impact, cost-effectiveness.

**WFP (2000) *Emergency Food Aid to Flood Affected People in Anhui, Hubei, Hunan and Jiangxi Provinces of China : Full Report of the Evaluation of China EMOP 6045* (Rome: WFP) Carlioni, A. et al**

**Objective:** Assess project impact and draw lessons for future emergency operations.

**Focus:** Food aid, LRRD, management, coordination, beneficiaries, gender.

**Criteria:** Impact, appropriateness, effectiveness, relevance, coverage, cost-effectiveness.

**WFP (2000a) *Emergency Assistance for Vulnerable Groups : Full Report of the Evaluation of DPR Korea EMOPS 5959.00 and 5959.01* (Rome: WFP) Landart, G., Van der Veen, A. and A. Wilkinson**

**Objective:** Examine the achievements of the emergency operation against its stated relief and recovery objectives.

**Focus:** Food security, nutrition, health, agriculture, coordination, logistics, gender.

**Criteria:** Impact, effectiveness, efficiency, appropriateness, coverage, cost-effectiveness.

---

## **Kosovo Evaluations (Chapters 3 & 4)**

### **Action by Churches Together (ACT)**

**ACT (2000) *Report of an Evaluation of EUBK91* (Geneva: ACT) Kieffer, G., de Klerk, T. and T. Silkin**

**Objective:** Review the performance of ACT as a unified operational structure in its management and coordination of the response to the Kosovo crisis.

**Focus:** Shelter, food security, agricultural rehabilitation, social development, de-mining, watsan, coordination, management.

**Criteria:** Efficiency, impact, coherence, effectiveness, appropriateness.

**Comments:** Covers Montenegro, Serbia and Kosovo.

### **Danish Refugee Council (DRC)**

**DRC (1999) *Lessons Learnt from the Danish Refugee Council's Involvement in the Kosovo Crisis 1998–1999* (Copenhagen: DRC) Paludan, A.**

**Objective:** Evaluate internal organisation of DRC response to crisis, its cooperation with partners and other actors, and the relevance of its intervention and approach.

**Focus:** Preparedness, organisational capacity, organisational structure, coordination, internal communication.

**Criteria:** Relevance, coherence, effectiveness.

### **Department For International Development, UK (DFID)**

**DFID (2000) *Coordination in the 1999 Kosovo Refugee Emergency : The Emergency Management Group (EMG), Albania* (London: DFID) Telford, J.**

**Objective:** Examine the strengths, weaknesses, role and capacity of the EMG as a coordination mechanism; produce recommendations for its future and lessons for other emergency responses.

**Focus:** Coordination, refugees, monitoring, management, LRRD.

**Criteria:** Effectiveness, efficiency, coherence, cost-effectiveness, coverage, relevance, appropriateness, sustainability, impact, adaptability.

### **Disasters Emergency Committee (DEC)**

**DEC (2000c) *Independent Evaluation of Expenditure of DEC Kosovo Appeal Fund: Phases I and II, April 1999–January 2000, Volumes I, II & III* (London: DEC) Wiles, P., Bradbury, M., Buchanan-Smith, M., Collins, S., Cosgrave, J., Hallam, A., Mece, M., Norman, N., Prodanovic, A., Shackman, J. and F. Watson**

**Objective:** Assess accountability to fundraising partners and British public, promote learning among DEC agencies, and monitor agencies' compliance to DEC rules and guidelines.

**Focus:** Food and nutrition, health, shelter, sanitation, education, mines, protection, psychosocial, coordination, war-affected populations and beneficiaries, LRRD, preparedness, performance standards.

**Criteria:** Appropriateness, coverage, effectiveness, impact, connectedness, coherence.

**Comments:** Volume III contains summaries of each member agency's DEC-funded activities.

### **European Community Humanitarian Office (ECHO)**

**ECHO (2000k) *Evaluation of the ECHO Actions in Kosovo Following the Kosovo Crisis: Final Report* (Brussels: ECHO) Alderson, M. J., Agger, I., de Boer, M. S. and E. Sondorp**

**Objective:** Review ECHO and partner response to the 1999 crisis following the NATO air strikes, and its coordination role in relation to UNHCR and others.

**Focus:** Rural and urban shelter, education, watsan, health, social and psychosocial protection, agriculture, de-mining, logistics, security, provision of non-food items, income generation, coordination, gender and strategy.

**Criteria:** Relevance, coherence, complementarity, effectiveness, efficiency, impact.

**Comments:** Contains summary covering Macedonia, Kosovo and Serbia of lessons learnt and recommendations, focusing on management issues and ECHO's relations with its partners and others.

***ECHO (2000l) Evaluation of the ECHO Actions in Serbia Following the Kosovo Crisis: Final Report (Brussels: ECHO)*** Alderson, J., Picard, E., Roeh, S. and E. Rogier

**Objective:** Review ECHO and partner response following the NATO air strikes during March–June 1999.

**Focus:** Social vulnerability, social protection, health, watsan, organisational issues, conflict prevention, refugees, IDPs, coordination, policy, management, monitoring.

**Criteria:** Relevance, efficiency, effectiveness, impact.

**Comments:** Includes annex on refugee return to Bosnia and Herzegovina and Croatia and summary of lessons learnt and recommendations covering Macedonia, Kosovo and Serbia focusing on ECHO management issues and its relations with partners and other actors.

***ECHO (2000m) Evaluation of the ECHO Actions in the Former Yugoslav Republic of Macedonia Following the Kosovo Crisis: Final Report (Brussels: ECHO)*** Alderson, J., Picard, E., Roeh, S. and E. Rogier

**Objective:** Review the response of ECHO and its partners to the humanitarian emergency in Macedonia arising from the influx of refugees fleeing the 1999 crisis.

**Focus:** Refugees and host families, social protection, health, conflict prevention, preparedness, coordination with partners and NATO military, rapid response, security, monitoring, LRRD, organisational activities.

**Criteria:** Relevance, effectiveness, efficiency, cost-effectiveness, impact, sustainability, complementarity.

**Comments:** Includes a cross-country summary, covering Macedonia, Kosovo and Serbia, of key lessons learnt and recommendations, focusing on ECHO management issues and its relations with partners and other actors.

***ECHO (2000n) Evaluation of ECHO Action in the Republic of Montenegro Following the Kosovo Crisis: Final Report (Brussels: ECHO)*** Capouillez, C., Kieffer, G. and M. Van Bruaene

**Objective:** Determine the strengths and weaknesses of ECHO's action in Montenegro following the 1999 crisis.

**Focus:** Refugees, IDPs, shelter, food and non-food distribution, health, psychosocial, preparedness, security, visibility, self-reliance projects, LRRD, coordination, the Roma.

**Criteria:** Relevance, efficiency, effectiveness, impact, sustainability.

**Comments:** Part of a series of evaluations examining the response of ECHO and its partners to the Kosovo crisis across the Balkan region.

***ECHO (2000o) Synthesis of the Evaluations of ECHO Actions Following the Kosovo Crisis (Brussels: ECHO)*** De Haulleville, A.

**Objective:** Synthesise the findings of five country evaluations and one thematic evaluation on coordination.

**Focus:** Preparedness, coordination, refugees, IDPs, policy, LRRD.

**Criteria:** Relevance, efficiency, effectiveness, impact, sustainability.

***ECHO (2000p) Evaluation of ECHO Action in the Republic of Albania Following the Kosovo Crisis: Final Report (Brussels: ECHO)*** Jadavji, S., Van Bruaene, M. J. and J. Welch

**Objective:** Assess the impact of action by ECHO and its partners since 1998 in Albania, and make recommendations for an exit strategy.

**Focus:** Public health, watsan, rehabilitation, social assistance, coordination, LRRD, exit strategy, response capacity, relations with military.

**Criteria:** Relevance, efficiency, effectiveness, impact, sustainability, cost-effectiveness.

**Comments:** Part of a series of evaluations examining the response of ECHO and its partners to the Kosovo crisis across the Balkan region.

**ECHO (2000q) *Evaluation of the Coordination between ECHO, Political Authorities, Military Forces and other Donors during the Kosovo Crisis: Final Report* (Brussels: ECHO) Kieffer, G., Paoli, C. and M. Van Bruaene**

**Objective:** Analyse ECHO's coordination with key international actors at the level of emergency assistance and transition and long-term LRRD.

**Focus:** Coordination, security, policy, protection, human rights, minorities and enclaves, LRRD.

**Criteria:** Coherence, connectedness, effectiveness.

**Comments:** Part of a series of evaluations examining the response of ECHO and its partners to the Kosovo crisis across the Balkan region.

## **International Committee of the Red Cross (ICRC) / International Federation of Red Cross and Red Crescent Societies (IFRC)**

**ICRC/IFRC (2000) *The Balkans Evaluation: An Examination of the Role of the International Red Cross and Red Crescent Movement's Response to the Balkans Crisis* (Geneva: ICRC) Anema, F., Stone, M. and H. Wissink**

**Objective:** Review the performance of the Red Cross and Red Crescent movement's integrated response to the Balkans crisis and invest lessons learned into future crisis management situations.

**Focus:** Coordination, management, beneficiaries, preparedness, human resources, communication, logistics, monitoring, ICRC and Federation integration.

**Criteria:** Effectiveness, efficiency, appropriateness, relevance, coherence, impact.

## **MSF-Holland (MSF-H)**

**MSF (1999c) *MSF Response in Macedonia to the Kosovo Refugee Crisis: A New Humanitarian Order?* (Amsterdam: MSF-H) Schulte-Hillen, C.**

**Objective:** Assess the effects of MSF-H's policy and NATO military action on the efficiency of MSF-H's interventions during the first two months of the Kosovo refugee crisis.

**Focus:** Policy, preparedness, refugees, humanitarian principles, financing, coordination, health, management.

**Criteria:** Effectiveness, efficiency, appropriateness, coverage.

## **Tearfund**

**Tearfund (2000) *Tearfund Balkans Emergency Response April 1999–July 2000 Evaluation* (London: Tearfund) Bakewell, O., Hume, W., Lavy, R. and C. Piper**

**Objective:** Evaluate the appropriateness and effectiveness of Tearfund International's response to the Balkans crisis, March 1999–May 2000.

**Focus:** Strategy, internal organisation, coordination.

**Criteria:** Appropriateness, efficiency, effectiveness, impact, complementarity, sustainability, coherence.

**Comments:** Covers Tearfund and partner's humanitarian action in Albania, Macedonia, Kosovo and FRY.

## **United Nations Children's Fund (UNICEF) / UK Department For International Development (DFID)**

**UNICEF/DFID (2000) *UNICEF Preparedness and Response in the 1999 Kosovo Refugee Emergency: A Joint UNICEF/DFID Evaluation* (New York: DFID/UNICEF) Greene, M., Madi, K., Stevens, R. and J. Telford**

**Objective:** Evaluate the strengths and weaknesses of UNICEF's preparedness and response activities in Kosovo, Albania and Macedonia.

**Focus:** Mine awareness, psychosocial, education, health, nutrition, operational support, capacity, child-friendly spaces.

**Criteria:** Appropriateness, impact, sustainability, cost-effectiveness, coordination, efficiency.

**Comments:** Includes lessons for DFID about evaluation in emergencies.

## **United Nations High Commission for Refugees (UNHCR)**

**UNHCR (2000a) *The Kosovo Refugee Crisis: An independent evaluation of UNHCR's emergency preparedness and response* (Geneva: UNHCR). Garlock, R., Barutciski, M., Sandison, P. and A. Suhrke**

**Objective:** Review a range of policy, operational and management issues focusing on UNHCR's preparedness in the year leading up to the crisis and its response following the March 1999 exodus.

**Focus:** Preparedness, relief, coordination, protection, management, relations with military, security, policy, refugees.

**Criteria:** Appropriateness, relevance, effectiveness, efficiency, coverage, coherence.

## **World Food Programme (WFP)**

**WFP (2000b) *Full Report of the Evaluation of the Kosovo Regional Emergency Operation* (Rome: WFP)**

**Objective:** Assess WFP's preparedness in 1998/99, its response to the crisis, and the effectiveness of its intervention and coordination mechanisms.

**Focus:** Preparedness, coordination, food aid, refugees, IDPs, human resources, communication, military.

**Criteria:** Effectiveness, efficiency, appropriateness.

## ANNEX 5

### REFERENCES FOR OTHER SOURCE MATERIAL

- Apthorpe, R. (1997) 'Some Relief from Development: Humanitarian Emergency Aid in the Horn of Africa (including Sudan), Rwanda and Liberia' in *European Journal of Development Research* vol. 9 no 2,83–106
- Apthorpe, R. and P. Atkinson (1999) *Towards Shared Social Learning for Humanitarian Programmes: A Synthesis Study* (London: ALNAP)
- Apthorpe, R. (2000) *Kosovo Humanitarian Programme Evaluations: Towards Synthesis, Meta-Analysis and Sixteen Propositions for Discussion*. Background Paper for the ALNAP Symposium 'Learning from Evaluation: Humanitarian Assistance and Protection in Kosovo' 17<sup>th</sup> & 18<sup>th</sup> October 2000, Geneva (London: ALNAP)
- Barutciski, M. (2000) *Solutions and Responses to the Humanitarian Crisis in Kosovo (1998-1999): Prevention and Preparedness Regarding the Refugee Emergency* (London: DFID/ESCOR)
- Borton, J. and J. Macrae (1997) *Evaluation Synthesis of Emergency Aid Evaluation Report 613* (London: DFID)
- Bouman, D. and T. Silkin (1999) *One, Two, Many... but not yet whole: Mid-term Evaluation Emergency Appeal EUKC91 of ACT in Macedonia and Albania* (Geneva: ACT)
- British Red Cross Society *Kosovo Crisis – IWD Response, Achievements, Challenges: What next?* (London: BRCS, Internal)
- CAFOD (1999) *Kosovo Emergency Programmes: Lessons Learning Workshop Report* (Tirana: CAFOD Albania)
- CAFOD (2000) *Albania Emergency Preparedness Workshop Report* (London: CAFOD)
- CARE International (2000) *The Kosovo Crisis – What can we learn?* (London: CARE International)
- CIDA (1997) *WID/Gender Equity Performance Review – Best Practices Study* (Hull: CIDA)
- Cliffe, L. and R. Luckham (2000) 'What Happens to the State in Conflict: political analysis as a tool for planning humanitarian assistance' in *Disasters*, Vol. 2 no 4, 291–313
- Corrin, C. (2000) *Gender Audit of Reconstruction Programmes in South Eastern Europe* (New York: Women's Commission for Refugee Women and Children, Urgent Action Fund)
- Daalder, I. and M. O'Hanlon (1999) 'Unlearning the Lessons of Kosovo' in *Foreign Policy* (Fall) (Washington, DC: Brookings Institute)
- Daalder, I. (1999a) 'Lessons of Kosovo' in *Die Woche* (June 24) (Washington, DC: Brookings Institute)
- Davidson, S. and P. Raynard (2001) *Ahead of the Field: Pilot Agencies and the People in Aid Code 1997–2000* (London: PIA)
- de Putte, B. (2001) *Follow-up to Evaluations of Humanitarian Programmes* (London: ALNAP)
- de Waal, A. (1997) *Famine Crimes: Politics and the Disaster Relief Industry in Africa* (Oxford: James Currey)
- Edwards, M. and D. Hulme (eds) (1995) *Non-Governmental Organisations – Performance and Accountability* (London: Earthscan/Save the Children)
- ETC UK (1999) *Real-Time Evaluation of the Humanitarian Response to the Crisis in Kosova, March to May 1999* (Copenhagen: DANIDA)

- Forman, S. and S. Patrick (eds) (2000) *Good Intentions: Pledges of Aid for Post-Conflict Recovery* (Boulder: Lynne Rienner)
- Groupe URD (1999) *Civilian Humanitarian and Military Actors in Albania* (France: Groupe URD)
- Groupe URD (2000) *8 Mois Apres Mitch: Bilan des Actions et Premieres Lecons* (France, Groupe URD)
- Groupe URD (2000a) *La Mouvance Associative Albanaise, la Crise Kosovar et les Strategies des ONG Francaises: Document de Synthese* (France: Groupe URD)
- Gunn, S. (1990) *Multilingual Dictionary of Disaster Medicine and International Relief* (Dordrecht: Kluwer Academic)
- Haas, R. (1999) 'The Lessons of Kosovo's Two Wars' in *IntellectualCapital.com* (June 10) (Washington, DC: Brookings Institute)
- Hallam, A. (1998) *Evaluating Humanitarian Assistance Programmes in Complex Emergencies. RRN Good Practice Review* (London: ODI)
- HPN (2000) *NGO Responses to Hurricane Mitch: Evaluations for Accountability and Learning. Humanitarian Network Paper 34* (London: ODI)
- IASC (2000) *Global Humanitarian Assistance 2000* (Geneva: UNOCHA)
- ICRC/IFRC (2000) *Technical Workshop on the Health & Relief Lessons Learned in the 1999 Balkans Crisis* Outcomes of a participatory workshop held 6–8 July 2000 at the Chateau de Penthes in Geneva for Member Societies of the Red Cross and Red Crescent Movement
- ICRC *International Humanitarian Law and the Kosovo Crisis: Lessons Learned or to be Learned* (Internal Review of the Red Cross No 837)
- ICRC *Conflict in the Balkans: Human Tragedies and the Challenge to Independent Humanitarian Action* (ICRC)
- ICVA (2000) 'Special Issue: the Independent Evaluation of UNHCR's Emergency Preparedness and Response to the Kosovo Refugee Crisis' in *TALKBACK: The Newsletter of the International Council of Voluntary Agencies (ICVA)* Vol. 2, Issue 1 (February)
- IFRC (2000) *World Disasters Report* (Geneva: IFRC)
- International Crisis Group (1999a) *War in the Balkans: Consequences of the Kosovo Conflict and Future Options for Kosovo and the Region (ICG Balkans Report No 61)* (Sarajevo: ICG)
- International Crisis Group (1999b) *Kosovo: Let's Learn from Bosnia: Models and Methods of International Administration (ICG Balkans Report No 66)* (Sarajevo: ICG)
- International Crisis Group (2000) *Kosovo Report Card (ICG Balkans Report No 100)* (Pristina/Brussels: ICG)
- JEEAR (1996) *The International Response to Conflict and Genocide: Lessons from the Rwanda Experience Vols. 1–4* (Copenhagen: The Steering Committee of the Joint Evaluation of Emergency Assistance to Rwanda)
- Joint Committee (1994) *The Program Evaluation Standards* (Thousand Oaks: Sage, Joint Committee on Standards for Educational Evaluation)
- Kirkpatrick, I. and M. Martinez Lucio (1995) *The Politics of Quality in the Public Sector* (Routledge: London)
- Leader, N. (2000) *The Politics of Principle: the Principles of Humanitarian Action in Practice. HPG Report 2* (London: ODI)

- Le Billon, P. (2000) *The Political Economy of War: What Relief Agencies Need to Know*. HPN Network Paper 33 (London: ODI)
- Lister, S. (2000) *Scaling-up After Hurricane Mitch: A Lesson Learning Study for the Disasters Emergency Committee* (London: ODI)
- Morris, N. (1999) 'UNHCR and Kosovo: a personal view from within UNHCR' in *Forced Migration Review* vol. 5 (August)
- Minear, L., Sommers, M. and T. van Baarda (2000) *NATO and Humanitarian Action in the Kosovo Crisis: Occasional Paper #36* (Providence, RI: Thomas J. Watson Jr. Institute for International Studies, Brown University)
- NATO (2000) *Kosovo One Year On: Achievements and Challenges*
- NATO/EAPC (1999) *Compendium of Views and Experiences on the Humanitarian Aspects of Peacekeeping* (Brussels: NATO)
- NATO/EAPC (1999) *Seminar on the Kosovo Experience with Regard to the Compendium of Views and Experiences on the Humanitarian Aspects of Peacekeeping* (Brussels: Euro-Atlantic Partnership Council)
- The Netherlands Government (2000) *Evaluation of the Kosovo Crisis* (The Netherlands)
- OECD-DAC (1999) *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies* (Paris: OECD)
- OECD-DAC (2000) *Glossary of Evaluation and Results Based Management (RBM) Terms* (Paris: OECD Working Party on Aid Evaluation)
- Ostrom, E. (1990) *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press)
- Oxfam Humanitarian Response to Kosovo Refugee Crisis (OXFAM)
- Oxfam International (2000) *Winning the Peace? Some Lessons from the EU's Aid to South Eastern Europe* (London: OXFAM)
- Patton, M. (1997) *Utilization-Focused Evaluation: The New Century Text* (Thousand Oaks: Sage)
- Paul, D. (1999) *Protection in Practice: Field Level Strategies for Protecting Civilians from Deliberate Harm*. RRN Network Paper 30 (London: ODI)
- Paul, D. (2001) *Protecting Rights through Humanitarian Action: Reference Group on Human Rights and Humanitarian Action* (New York, IASC)
- Perrin, R. (1999) *Presentation to the UK Evaluation Society's Annual Conference* (December)
- Pictet, J. (1958) *The Geneva Conventions of 12 August 1949: Commentary*. Published under the general editorship of Jean S Pictet, Director for General Affairs of the International Committee of the Red Cross (Geneva: ICRC)
- Porter, T. (1999) 'Coordination in the Midst of Chaos: the Refugee Crisis in Albania' in *Forced Migration Review* vol. 5 (August)
- Pottier, J. (1999) 'In Retrospect: Beneficiary Surveys in Rwandan Refugee Camps, 1995: Reflections 1999', in Frerks, G. and T. Hilhorst (eds.) *Evaluating Humanitarian Aid: Politics, Perspectives and Practice* (Wageningen: Disasters Studies Unit, University of Wageningen) 115–126
- Raynard, P. (2000) *Mapping Accountability in Humanitarian Assistance* (London: ALNAP)
- RDI (1990) *Evaluation of ODA Support for NGO Post-Flood Rehabilitation Activities in Bangladesh 1988–89* (London: Relief and Development Institute)

- Refugee Studies Programme (1999) *Forced Migration Review* vol. 5 (August)
- Scriven, M. (1991) *Evaluation Thesaurus – Fourth Edition* (Newbury Park, CA: Sage)
- Sobhan, R. (1997) 'Rethinking Technical Assistance' in *Discourse: a Journal of Policy Studies*, vol. 1 no 1, 1–16
- Spencer, T. (1998) *A Synthesis of Evaluations of Peacebuilding Activities Undertaken by Humanitarian Agencies and Conflict Resolution Organisations* (London: ALNAP)
- SPHERE (2000) *Humanitarian Charter and Minimum Standards in Disaster Response* (Geneva: The Sphere Project)
- Stufflebeam, D. (1999) *Foundational Models for 21<sup>st</sup> Century Program Evaluation* (Western Michigan University: The Evaluation Centre Occasional Papers Series)
- Suhrke, A. (2000) 'From One Crisis to Another: Organisational Learning in UNHCR' in *Learning in Development Cooperation*, Wohlgemuth, L. and J. Carlsson (eds) (Stockholm: Expert Group on Development Issues, EGDI)
- Tiller, S. (2000) *Kosovo Schools Rehabilitation Programme Review* (London: British Red Cross Society)
- Valid International (1999) *Kosovo Crisis: Lesson Learning Study* (London: DEC)
- UK House of Commons Select Committee on International Development (1999) *Kosovo: The Humanitarian Crisis* (London: International Development Committee)
- UK National Audit Office (2000) *DFID Emergency Aid: The Kosovo Crisis* (London: The Stationary Office)
- UK Ministry of Defence (1999) *Kosovo: Lessons from the Crisis* (Internal Report)
- UNDP (2000a) *Sustainable Human Development Good Project Design: An Assessment of Project Formulation and Design in UNDP Agriculture Programming, 1994–1999* (New York: UNDP)
- United Nations (2000) *Consolidated Inter-Agency Appeal for Southeastern Europe: Mid-Term Review* (Geneva: UNOCHA)
- UNTAET (2000a) *East Timor Consolidated Appeal Process (CAP) Review Phase 1. Self-Assessment by Humanitarian Community* (Dili: Consolidated Inter-Agency Appeal Process Steering Committee, UNTAET Humanitarian Assistance and Emergency Rehabilitation Pillar)
- UNTAET (2000b) *East Timor Consolidated Appeal Process (CAP) Review Phase 2. Humanitarian Programs in East Timor from the Beneficiaries' Perspective* (Dili: UNTAET)
- USAID Center for Development Information and Evaluation (1997) 'The Role of Evaluation in USAID' in *Performance Monitoring and Evaluation Tips No.11* ([http://www.dec.org/pdf\\_docs/pnaby239.pdf](http://www.dec.org/pdf_docs/pnaby239.pdf))
- USAID Center for Development Information and Evaluation (1999) *Impact Evaluation: Providing Emergency Aid to Haiti* (Washington DC: USAID)
- US Department of Defence (2000) *Report to Congress: Kosovo/Operation Allied Force After Action Report* (Washington: US Department of Defence)
- US Department of State (2000) *Interagency Review of U.S. Government Civilian Humanitarian and Transition Programs* (Washington: The National Security Archive)
- Valadez, J. and M. Bamberger (1994) *Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers, and Researchers* (Washington DC: World Bank EDI Development Series)

- Valid International (1999) *Disasters Emergency Committee Kosovo Crisis: Lesson Learning Study* (London: DEC)
- Van Brabant, K. (1999) *Opening the Black Box: An Outline of a Framework to Understand, Promote and Evaluate Humanitarian Coordination* (London: ODI)
- Van Brabant, K. (2000) *Operational Security Management in Violent Environments: A Field Manual for Aid Agencies* Good Practice Review 8 (London: ODI)
- Von Flüe, C. and J. De Maio (1999) *Third Workshop on Protection for Human Rights and Humanitarian Organisations* Report on the Workshop, Geneva, 18–20 January 1999 (Geneva: ICRC)
- WFP Executive Board (2000a) *Summary Evaluation Report on Kosovo Regional Emergency Operation 6040: Food Assistance to Kosovar Internally Displaced Persons and Refugees in the Federal Republic of Yugoslavia and the Former Yugoslav Republic of Macedonia* (Geneva: WFP)
- WFP Executive Board (2000b) *Information Notes: Summary of Evaluation Recommendations and Management Response – Kosovo Regional Emergency Operation 6040* (Geneva: WFP)
- Wholey, J. (1997) 'Trends in Performance Measurement: Challenges for Evaluators' in Chelimsky, E. and W. R. Shadish (eds.) *Evaluation for the 21<sup>st</sup> Century: A Handbook* (Thousand Oaks: Sage)
- Wood, A., Borton, J. and R. Apthorpe (eds.) (2001) *Evaluating International Humanitarian Action: Perspectives from Practitioners* (London: Zed Press)
- Works, B. (1999) *Preliminary Report: Kosovo & Operation Allied Force – Lessons to be Learned* (Washington DC: Strategic Issues Research Institute; SIRI-US)
- WHO Inter-agency Medical/Health Taskforce: *Humanitarian Crisis in the South Balkans* (WHO Guidance Document)