

the context of one bilateral agency (DANIDA) being over-represented, and others (JICA, USAID, GTZ) being comparatively 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 p58). While relatively strong on contextual background, Terms of Reference, conclusions and recommendations and legibility, evaluation reports are substantively insufficient, particularly with regard to clarity in presentation of methodology and the methodology itself, attention to international standards, and attention to gender. Weakness in these substantive 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, the terms of which need to be enforced. Furthermore, findings may 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, diarrhea outbreaks occurred, but were effectively controlled by adequate

#### **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 health care 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)

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).

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 proven 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, for example, ECHO, 1999, 2000a; MSF, 1999).

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 contextualise '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

## 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.’

(DEC, 2000b: 12.9.8)

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.

As the literature makes clear, 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 contextualisation 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 humanitarian 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 p46 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 to the complex emergencies and natural disasters to which they were responding, as well as 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 the affected population’s perspective. Because of the apparent lack of adequate consultation with beneficiaries and non-beneficiaries by 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) still 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 cross-checking, for as DEC perhaps ironically comments: ‘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, 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; also see 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 in attempting to achieve a similar objective. 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; also see 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 '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) that 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 most evaluators and commissioning agencies are not 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, or an inability to operationalise, the standards. 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 (for example, in particular the DEC evaluation reports, but also UNTAET) required a focus in this area.

Some evaluation reports provide a limited, indeed tantalising, amount of information about coping strategies, but without systematic discussion. Examples of coping strategies from the emergency phase are provided in Box 2.3. 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 remain in their environments 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 concerns psychological and emotional needs. An exception here is the DEC (2000), which notes that there was scant evidence that psychological and emotional needs were properly identified and considered in responses. Another important exception is that of the Norwegian Ministry of Foreign Affairs (1999). This thorough and comprehensive evaluation report and planning document

**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 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)

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).

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 of its area of expertise to housing rehabilitation caused various problems such as lack of appropriate siting and use of inappropriate materials;

**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 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 Columbia, is provided in Box 2.5.

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

**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)

LRRD Debate for Establishing Evaluation Standards (Chapter 4, p132).

### ***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 the longer 1992–1998 period.

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 authorised 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 DANIDA (1999h) comments, partly as a result of the failure to provide protection there is now a greater readiness to discuss conditionality on humanitarian assistance to attempt to ensure that affected populations are protected, 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***

Lack of preparedness by governments and agencies is a common finding in the evaluation reports that consider this issue. 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 fully-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.

- 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). Further details on M&E, and recommendations for strengthening 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 around 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 is recommended by a majority of evaluation reports, including coordination with national governments
- 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 through the lens of a proforma that has been used in this and 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 the objective – as requested by ALNAP Full Members – 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) The standards developed by the Joint Committee on Standards for Educational Evaluation (Joint Committee, 1994) were also used (see Chapter 4 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 around the measure of 'adequacy'.

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 Experienced***

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