Australian Agency for International Development

AusAID

Managing Emergency Humanitarian Aid Evaluation: Lessons from Experience

Evaluation No: 12 - March 1998

Australian Agency for International Development

Managing Emergency Humanitarian Aid Evaluation: Lessons from Experience

Raymond Apthorpe
Ann Nevile

Evaluation No: 12 - March 1998

Managing Emergency Humanitarian Aid Evaluation: Lessons from Experience
Evaluation No: 12
March 1998

ISBN 0 642 22070 0
ISSN 1030-7249

A report on the main findings of a workshop held with AusAID support at the National Centre for Development Studies, Australian National University, Canberra 13 - 16 March 1998

Printed by CPN Canberra 1998

Contents

Preface iv
Executive Summary v
Section 1: Emergency Humanitarian Aid 1
Natural disasters 2
Complex emergencies 2
Section 2: Ways of Thinking About Evaluation 5
Logical frameworks 7
Cost-effectiveness reviews 7
Institutional analysis 7
Evaluation as negotiation 8
Section 3: What Should be Evaluated? 9
Terms of reference 10
Who should draw up the terms of reference? 10
Selecting team members 12
Section 4: What Should Come Out of the Evaluation? 15
Discussing findings and conclusions 15
Making recommendations 17
Follow-up action 18
References 20
Annex 1: List of Participants 22
Annex 2: Workshop Agenda 24

Preface

A workshop, jointly organised by Raymond Apthorpe at the National Centre for Development Studies (NCDS), Australian National University and John Borton at the Overseas Development Institute (ODI), London, brought together evaluation specialists and practitioners with experience in evaluating complex emergencies.[1] The workshop was funded by AusAID and was held over four days (13 - 16 March 1998) at NCDS, Australian National University, Canberra. This report summarises workshop discussions and was distributed at the May 1998 meeting of the Development Assistance Committee (DAC)'s Working Party on Aid Evaluation.

As well as providing the basis for this report, the workshop provided an opportunity for input into the following: ODI's work in progress on the DAC's Good Practice in Evaluating Humanitarian Aid initiative; a collection of papers discussing four major emergency humanitarian aid evaluations (to be published in the journal *Disasters*); a proposed set of case studies on logical frameworks and their relevance to humanitarian aid; and a proposed set of case studies focusing on the beneficiaries of humanitarian emergency aid.

1 Our thanks for administrative help go particularly to our colleagues at our respective institutions: Helen Awan, Jennie Colman, Liz Ingram and Maurette Macleod. For helpful comments on the text of this report we are grateful to John Borton and Katherine Ruiz-Avila.

Executive Summary

Emergency humanitarian aid is provided in times of disaster or crisis. The most complex emergencies are those which occur in areas where there is disputed sovereignty manifested in armed conflict. Complex emergencies are characterised by the collapse of social infrastructure, risks beyond everyday life, indiscriminate as well as discriminate violence, and the extreme material, organisational and psychological vulnerability of those caught up in the emergency situation. The difficulties inherent in the design and delivery of humanitarian assistance are exacerbated by the need to operate in such an unstable, turbulent and insecure environment.

Recognition of the difficulties involved in providing humanitarian assistance in such an environment, together with an increasing demand for emergency assistance at a time when aid expenditure (in real terms) is falling, has led many donors to ask questions about the appropriateness and effectiveness of emergency humanitarian aid. The evaluation of emergency humanitarian aid is therefore becoming increasingly important. This report attempts to address some of the hurdles faced in undertaking evaluations of complex emergencies, and to suggest solutions for overcoming them.

Evaluation of emergency humanitarian aid usually focuses on project operations. There is a case for large, system-wide evaluations to also address policy issues. This proposal is not without difficulties, since policy goals or objectives may not be clearly defined, and they may change in response to changing circumstances on the ground. This is particularly true of complex emergencies which are, by their nature, unstable and often protracted, stem from political uncertainties, and are conflict-driven. Evaluations of complex emergencies should evaluate the link between the core problem (conflict), program objectives and project activities.

The very nature of the environment in which emergency humanitarian assistance is provided suggests that evaluators of this assistance must adopt a flexible approach and be prepared to use a number of different methods. Traditional evaluation techniques of impact, efficiency and cost-effectiveness must be supplemented with techniques which allow the evaluator of humanitarian aid to focus on the actions and intentions of a diverse range of actors. Where possible, a degree of flexibility should be built into terms of reference (TOR) which allow evaluators to respond to unanticipated

questions and issues as they arise during the course of the evaluation. Another crucial element in the evaluation process is the selection of a multi-disciplinary evaluation team. A search process for experienced individuals should be combined with a tender process.

Evaluation reports need to provide recommendations based on their findings to improve the international response to future complex emergencies. To what extent these recommendations should be specific and targeted is a matter of debate. Large, system-wide evaluations raise issues relevant to a diverse range of organisations, and compliance with recommendations cannot be compelled. Monitoring of follow-up action can serve to increase the accountability of agencies to learn from the lessons of the past.

Section 1

Emergency Humanitarian Aid

Emergency humanitarian aid is, in many ways, the main public face of aid. Those outside the aid community can easily understand the need for assistance in response to a disaster or crisis that threatens human life. Needs seem clear and the appropriate response also seems clear and straightforward - provide food, water, shelter and basic health needs until the disaster or crisis situation is over, when, by implication, life returns to "normal". However, those involved in the delivery of emergency humanitarian aid are well aware that this simple equation belies the complexities involved in delivering effective humanitarian aid.

Emergency humanitarian aid is provided in times of disaster or crisis. Those providing the aid (both donor and implementing agencies) must respond quickly. The disaster or crisis situation may change rapidly. There is a need to balance short and long-term objectives when alleviating immediate suffering in ways which will not cause longer term damage to those caught up in the disaster or emergency, as well as to local economic and social structures. For example, provision of emergency food aid may undermine the economic viability of local farm production and may therefore prolong the humanitarian crisis. Similarly, while it may be easier to provide humanitarian aid to refugees and internally displaced people than to those who remain in their home areas, fleeing populations are rarely able to take with them the means of livelihood support and therefore the need for emergency humanitarian assistance is prolonged (Kirkby, 1998:6).

Furthermore, disasters or crisis situations are not homogenous. The appropriate response to a "natural" disaster, such as a drought or an earthquake, will almost certainly be quite different from that required in a situation where there is disputed sovereignty and armed conflict. To be effective, "the delivery of emergency humanitarian aid must always be related specifically to the cultural, political, military and environmental reality of the situation and to the capacity at the time of the delivery " (O'Keefe et al, 1997:2).

Natural disasters

Natural disasters are critical disruptions of everyday life by geo-physical hazards such as drought, flood, frost and earthquakes. Of course the intensity of disruption is affected also by pre-existing vulnerabilities, poverties, and inappropriate public policies and their effects on people's livelihoods. Therefore, the convenience of the man-made: natural classification amounts to a misnomer where the effects of disasters are seen to be caused as much or more by the social conditions they shake as by the geo-physical agents that precipitate them (UNDRO:UNEP, 1986, Quarantelli, discussed in Alexander, 1997).

Natural disasters still tend also to be contrasted with "complex emergencies" despite the similarities of major factors. "Complexity" is also true of the structural inter-meshing with situational causes of the intensity and duration of recurring famines in Ethiopia, or North Korea today, as of war (cf. Borton, 1994). Indeed, historically in Africa and elsewhere, famine and war have for centuries been closely inter-related.

Complex emergencies

Complex emergencies are conflict-driven disruptions characterised by a multiplicity of actors, resources, goals and objectives. The multiplicity of actors encompasses national governments, UN agencies, medical NGOs, NGOs involved in the delivery of food aid, local troops, armed civilians, local power figures, and intended and unintended beneficiaries. While there are incentives for local power figures and agencies to work together - the local power figures are able to win material benefits from the agencies and the agencies are seen to be "delivering aid" - goals may not always converge.

For example, in managing refugee settlements near the Kenyan/Somali border in 1991, UNHCR had to balance a number of diverse interests and pressures. The camps initially provided a haven for those fleeing armed conflict, but they also attracted refugees from Somalia who were not displaced by war, as well as Kenyan nationals wishing to access repatriation benefits. Similarly, the goal of the Government of Kenya - rapid repatriation of the refugee population - was not shared by UN agencies. UNOSOM, for example, was advising against repatriation because of the continuing insecurity in large areas of Somalia (Kirkby, Kliest & O'Keefe, 1997:15).

The differing perceptions of agencies and recipients, both of themselves and the other party, is a clear illustration of how key actors do not necessarily share the same value system, nor agree on "the rules of the game". For those involved in the delivery (and evaluation) of complex emergencies, it is therefore important to find ways of taking into account the perspectives of intended beneficiaries.[2] However, as yet, there is little information available on beneficiary perspectives and research on this is needed.

While a multiplicity of actors, resources and goals characterise many types of emergency situations, complex emergencies occur in areas where there is disputed sovereignty manifested in armed conflict. Complex emergencies are, therefore, characterised by the collapse of social infrastructure, risks beyond everyday life, indiscriminate as well as discriminate violence, and the extreme material, organisational and psychological vulnerability of those caught up in the emergency situation. The difficulties inherent in the design and delivery of (and consequently, the evaluation of) humanitarian assistance are exacerbated by the need to operate in such an unstable, turbulent and insecure environment.

The level of uncertainty and unpredictability inherent in complex emergencies means that plans have to be changed more frequently than is the case in other types of emergency situations. There are times in complex emergencies when the speed of the response is critical, and short-term/long-term trade-offs are significant as in the case of mass population movements.[3] What first appears as a creative solution to a problem can have unanticipated and unintended consequences. For example, UNHCR overcame the problem of how to provide security in refugee camps in eastern Zaire - Western governments, with the exception of the Netherlands, were unwilling to provide a security force - by using the local military (Adelman, 1998:4). Yet the *Joint Evaluation of Emergency Assistance to Rwanda: A Review of Follow-up and Impact Fifteen Months after Publication* noted that it was the presence of the militia in the camps, and the failure to deal with them, which was the catalyst to

2 "Intended beneficiaries are those in bona fide need who have been successfully identified and provided with aid... Unintended beneficiaries are those who gain illicitly through theft, corruption, misappropriation, or, as in Somalia, through demanding a payment for an agency's right to transport materials" (Kirkby, 1998:4).

3 However, for most complex emergencies, typically the situation to be addressed is chronic (nearly three decades in the case of Eritrea and almost two decades in the case of Mozambique) , and it is possible to introduce annual programming of interventions in such cases.

the conflict in eastern Zaire and the major factor preventing the rehabilitation of refugees (Adelman, 1998:5). Also important in this case was a justified fear of persecution which prevented rehabilitation in the period before the (forced) mass repatriation on November 1996, and, in 1994, the fact that only Bangladesh was willing to contribute troops to sweep the camps (Borton, 1998:Box 4).

The disputed sovereignty which characterises complex emergencies means that implementing agencies are constantly having to negotiate with local power structures, which may themselves be in a state of flux, or, indeed, evading such negotiation. The study of UNHCR's Cross Border and Cross Mandate Operation in Somalia and Kenya noted that "negotiation with the relevant authorities, including negotiations actually to be allowed to work, was more crucial than even the specific technical and investment planning of the operation" (Kirkby, Kliest & O'Keefe, 1997:17).

The need for protection of agency staff and refugees - and now also evaluators[4] - is an on-going issue. Judgements about what value is placed on the life of an expatriate compared to the value placed on the life of a local staff member have become a significant determinant of the cost of providing humanitarian assistance in the midst of a complex emergency. Staff turnover is very high (in Somalia, agency staff worked for an average of three months at a time), leading to minimal corporate memory and poor record keeping. Truck drivers and convoy leaders are attacked. Records may be destroyed. All of this contributes to a sense of insecurity, turbulence and chaos.

Recognition of the difficulties involved in providing humanitarian assistance in such an environment, together with an increasing demand for emergency assistance at a time when aid expenditure (in real terms) is falling, has led to urgent demands for evaluations of the appropriateness and effectiveness of emergency humanitarian aid.

4 For example, selection of representative cases to be evaluated in one recent evaluation exercise, was on the basis of safety considerations for evaluators, as well as program issues.

Section 2

Ways of Thinking About Evaluation

Evaluating emergency humanitarian aid which is being provided in uncertain, turbulent and insecure environments obviously presents challenges beyond those encountered in the evaluation of structural development aid. How should evaluators approach their task? The very nature of the environment in which emergency humanitarian assistance is provided suggests that evaluators must adopt a flexible approach, and be prepared to use a number of different methods. But what analytical tools will assist evaluators in situations where "hard" base-line data is non-existent, there is little information on assumed underlying causal relationships, it is difficult to collect monitoring data and, probably, there is no clear framework of objectives, timetables or exit strategies?

Traditional social science research wants to explain things. But it also wants to connect cause and effect. It does this by separating the effect you are trying to explain from the possible cause(s) of that effect. However, when looking at complex, interdependent relationships, it is not always possible to separate cause and effect because the direction of influence is often circular rather than linear. For example, institutional structures affect the distribution of resources, which in turn affects the interests and power of political actors. But political actors may change the decision-making rules and formal institutional structures in order to change the distribution of resources and entrench their power. In other words, what has been identified as the cause (institutional structures) can be changed by the effect (the power of political actors).

Thus methods which are more common in historical and philosophical research are often more productive than those traditionally employed in the social sciences. Such methods acknowledge the complexity and interdependent nature of events in the real world and ask not, 'did *x* cause *y*?' but rather, 'what happened?' and 'why?' In other words, in order to understand and be able to deal with situations and structures, they seek to build narratives about specific events and processes, rather than theorising grandly and establishing causal relationships.

Evaluation has much to gain by borrowing from these historical and philosophical methods. The first step in using this approach to analyse a particular set of events and processes is to construct a narrative history to serve as a kind of model or spectrum through which to view what is to be understood and explained. You ask many different actors to tell you their story, recognising that what you are being told is the truth as they see it, or the truth as they would like you to see it. What a researcher hopes to gain from this particular approach is a partial understanding of someone else's view of reality (Reinharz, 1992:18). The stories of many different actors are then added to the available documentary evidence to construct the narrative history. The narrative history is more than a simple chronology. It details not just what happened and when, but also who was involved, and links significant events. For an evaluator, this narrative history establishes your baseline, or template, which helps you make judgements about events and processes.

The evaluator then has to go beyond this baseline, or template, and explain why actors did what they did and with what effects. In other words, filter the narrative (raw data) through some sort of analytical framework and draw practical conclusions accordingly. The framework may arise out of the stories themselves. When many different stories are accumulated, consistent patterns may emerge. Alternatively, the framework may arise out of previous studies and be confirmed by the evidence revealed in the narrative history.

At the same time, it must be remembered that consultancy is not research, nor research consultancy: each has its own standards and practices. Currently, most evaluations of humanitarian aid are done through consultancies. The main exception to this general rule is the Rwanda Joint evaluation. In this case, the level of resources and time committed was such that the research/consultancy line was significantly blurred.

The narrative history approach described above is particularly suited to evaluating complex emergencies because it relies on qualitative rather than quantitative data; because it allows the evaluator to focus on actions and intentions; and because it highlights the competing agendas of the diverse range of actors. Furthermore, use of the narrative history approach does not preclude the use of more analytical models, such as logical frameworks or cost-effectiveness reviews, to examine specific components of an emergency assistance program. For example, the large Rwanda evaluation supplemented traditional evaluation techniques of impact, efficiency and cost-

effectiveness analysis with "qualitative analysis of cause-and-effect assessed in relation to contractual obligations or international legal norms. Results had to be measured against political and ethical standards as well as international humanitarian legal and professional standards" (Dabelstein, 1996:291).

Logical frameworks

The logical framework has potential for evaluating humanitarian relief programs, emergencies, and complex emergencies. For example it can be used to improve planning and evaluation approaches that are fixated on shorter-term outputs (cf. Gasper, 1998). A number of modifications will be necessary, however, before what has not been devised particularly for evaluations of emergency aid can be used for this purpose. Especially if the objectives hierarchy is not very clear, the greater attention that ought to be paid to assumptions analysis in emergency aid evaluations (assumptions analysis tends to be marginalised in actual practice in structural aid evaluations) will be difficult. Furthermore, there are, as always with the use of this framework, issues to disentangle of applications as between project and program evaluation, regardless of the type or field of aid.

The Canberra meeting recommended that a research project in policy and project analysis should explore possible applications of logframe analysis to emergency relief.

Cost-effectiveness reviews

Cost-effectiveness has great potential for complex emergency aid ex ante appraisal, as well as evaluation looking for lessons to be learned (Hallam, 96). Efficiency is easier to measure than effectiveness. Cost-effectiveness, unlike cost-benefit analysis (but despite their similar objectives) can be used where analysts are either unable or unwilling to monetise benefits. Nonetheless, using CBA to assess cost-effectiveness remains an ambitious objective, especially where outcomes and impacts are hard to define and an evaluation of disaster-wide effectiveness is sought rather than interventions by individual agencies (Bradly, 1998).

Institutional analysis

Program analysis particularly, but operations analysis too, calls for institutional as well as other types of analysis. As we have seen, a multiplicity of organisations is involved in complex emergencies.

Institutional and organisational assessment is therefore all the more necessary - and sensitive and difficult. The skills required come mainly, for example, from training in social anthropology, law, political science, sociology and social psychology. Nevertheless, if too many economists are involved in structural aid evaluation, too few are involved in humanitarian aid evaluation, where the problems as well as the response to be assessed are motivated by economic as well as political and humanitarian concerns.

Institutional analysis in a political economy mode is particularly applicable to complex emergencies (cf. Donini et al, 1996). Agencies' concepts and definitions of "emergency" must be interpreted to see to what extent they reflect more their own organisational exigencies, including those of resources, or more those of the situations into which they have (or have not) intervened. Victims' and beneficiaries' understandings and responses must be evaluated much more than is the case in actual practice using consultation methods yet to be devised. As for channels and means of emergency aid distribution, complementarities, for instance as between charitable and commercial organisations, could be sought (cf. Kent, 1996).

Evaluation as Negotiation

Evaluation conceived more as a consultative process of negotiation, than intervention through packages of incidental sectoral expertise, has much to offer for example where everything - especially institutions and organisations - seems to be in flux. Numerous parties contend, outcomes are required that satisfy all, received development wisdoms seem not to apply (for example, past structural aid has not obviated the need for present relief aid), and multiple objectives rule the day. Complex more than natural emergencies are obvious candidates for such an approach to evaluation which, except perhaps in certain cases of self-evaluation and similar internal concerns, are uncommon in actual practice in outsider and mixed team evaluations.

Who participates, when and where, under what conditions, in what manner, against which resources, backgrounds and contexts, and with what expectations and rewards, normally are crucial to the success of negotiations. Evaluations conceived as forms of negotiation, including TOR which call for agreements to be reached with all parties as well as for specialist advice brought to bear on crucial issues, would depend for their success on similar factors.

Section 3

What Should be Evaluated?

Those planning an evaluation must first ask themselves whether a project (ie. operations) or program (ie. policy) evaluation is required. Evaluation of emergency humanitarian aid usually focuses on project operations. Donors feel more comfortable evaluating the projects through which a policy was implemented, rather than evaluating the assumptions which lie behind a particular policy. However, more recent system-wide humanitarian aid evaluations (such as the Rwanda study) have addressed policy issues. System-wide evaluations are impossible without sufficient consensus on the part of the many donors and agencies involved on the need for such an evaluation. In the case of Rwanda, the massive amount of money involved in relief operations, together with large-scale genocide, was sufficient to produce consensus within the donor community on the need for a system-wide evaluation. In this case, the consensus-building process was also facilitated by the various social, political and foreign policy objectives, as well as some informed intuition (if nothing else) that there remained some chapters of the story to run.

Where consensus on the need for a system-wide evaluation has been reached, the next question that arises is whether norms on which the policy is based should be evaluated? The answer is yes. Underlying norms should be evaluated if these norms are not the ones under which people were actually operating. For example, Study II of the Joint Rwanda evaluation asked, 'what are the stated norms?', 'what norms underlie actual practice?', and then examined discrepancies between stated norms and practice. If the hierarchy of objectives is revealed in actions rather than written documents or statements, new sensitivities arise, but also the opportunity for a richer account.

Policy goals or objectives are seldom clear or sharp enough to serve as criteria against which to test management and performance. They are not immediately evaluable in their own terms. In particular, those involved in policy evaluation must take the stated objectives of policies, projects and evaluation TOR, and translate them into something evaluable. Recognising that goals and objectives may change in response to changing circumstances on the ground is

another requirement for best evaluation practice. This is particularly true of complex emergencies, which are, by their nature, unstable, yet also very often protracted or recurring.

Terms of reference

The question posed at the beginning of this section (policy or project evaluation?) illustrates the general issue of needing to be precise about the purpose of an evaluation so that the TOR can highlight relevant issues. This need for precision requires those drawing up the TOR to go beyond a repetition of the effectiveness, efficiency and impact mantra. TOR should be grounded in the broader question of what is the problem and its aetiology that you are trying to address? For example, it is essential that ways be found for evaluations of complex emergencies to look at both humanitarian aid delivery and conflict management. In other words, evaluate the link between the core problems, which have resulted in conflict, and program objectives.

In research, an explicitly comparative element may help researchers get closer to the truth. But in consultancy, what do you compare? The performance of different donor agencies? Individual donor agencies would be reluctant to participate in any evaluation which explicitly compared performance. Similar problems emerge with explicit comparisons of implementing agencies. NGOs operate within a competitive market and would therefore see any explicit comparison as an exercise which could reduce their individual market share. Do you compare present practice with past performance? Do you ask whether the agency has learnt from the past? Do you include questions in the TOR which ask, 'to what extent has the agency evaluated its own programs?' 'to what extent has the agency carried out independent audits?' 'to what extent has the donor(s) required these sort of audits?' and 'to what extent has the donor agency incorporated previous evaluation findings into current policies and programs?' Asking such questions may not generate base-line data, but it may, over time, raise the awareness within donor agencies of the need to integrate evaluation into policy development, and signal to implementing agencies that donor agencies value self-evaluation (cf. Martin, 1998).

Who should draw up the terms of reference?

The issue of who should be involved in drawing up the TOR is related to questions of content. Different actors have different agendas and these agendas will inevitably influence the TOR.

Important as it is to identify the broad issues, TOR may also contain very specific questions. However, these specific questions must not be immutable. Sometimes it becomes clear halfway through an evaluation that the specific questions set out in the TOR are not appropriate and other questions should be asked. TOR must therefore be treated as negotiable by all parties at the earliest possible stage in the evaluation process (Apthorpe, 1998). Donor agencies must be equally flexible in their funding arrangements.[5] Flexible contracts were used in the Somalia evaluation which meant that additional resources could be called upon as needs emerged, thereby ensuring timely completion of the evaluation report. In this case, the whole exercise took twelve months from the initial preparatory work to publication of the final report.

The problems of dealing with flexible TOR may be reduced if a preliminary design phase, in which all parties participate, precedes the main contract. A preliminary design phase, or scoping study, allows the evaluator to identify potentially significant issues and prepare a more accurate schedule and budget for the remainder of the evaluation.

How far you can go in accommodating differing points of view is by no means fixed. In some situations it may be impossible to reconcile differing agendas. For example, in the case of the Rwanda evaluation, not all DAC members wanted the evaluation to include a focus on the conflict itself, as well as the emergency assistance program. It was therefore not possible to conduct the evaluation as a joint DAC exercise. This opened up the possibility of other players (UN agencies and NGOs) being involved in the steering committee. In this case, it was the steering committee, which drew up the TOR, approved team composition and commented on the draft report. The inclusive nature of the steering committee (the fact that a range of different types of agencies were represented on the committee and all were on an equal footing) was identified as one of the strengths of the Rwanda evaluation. However, membership of steering committees or management teams should not be restricted to individuals who work in head offices. Field officers and other local players should also be part of the management team.

Feelings of broader ownership may also be generated if the TOR are developed through a workshop process, in which insiders and

5 The need for flexibility on the part of donor agencies does not, however, excuse those undertaking the evaluation from the responsibility of making a realistic assessment of whether the TOR can be addressed adequately in the proposed time-frame.

outsiders participate. Staff then feel the evaluation will produce something of value to them.

Selecting team members

The Rwanda evaluation was characterised by good working relations between management (the steering committee) and the evaluation teams. The steering committee adopted a flexible approach to the TOR and there was a high degree of trust (and hence cooperation and information flows) between management and the evaluators, several of whom were also academic researchers. But was this high level of trust and cooperation a lucky accident? Obviously the individual characteristics of team members will have some effect, but the way in which a team is selected is also important.

Experience at large suggests that a tender process alone does not produce the optimum result. Good teams are more likely to be produced by a process of short-listing suitable individuals, discussing their strengths and weaknesses, and then asking individuals if they are interested, or confining a bidding process to those on the short-list. In other words, a search process should be combined with a tender process. Evaluations of humanitarian assistance in "natural disasters" have the advantage usually of a stable - or at least a continuous-political (and policy) environment. Nevertheless longer - as well as shorter-term time perspectives are needed, structural as well as situational analysis and a broad range of sectoral expertise.

When evaluating a complex emergency, a multi-disciplinary team is essential because of the range of issues involved. Optimum team composition will vary, however, depending on the particular characteristics of each disaster or emergency. The very nature of complex emergencies indicates that an evaluation team requires experts on conflict, someone who understands structural/institutional trends under stress as well as those equipped to understand the immediate survival needs of those caught up in the emergency. The strength of a multi-disciplinary team lies in the differing perspectives it can bring to bear on issues. Therefore, regardless of the exact structure of the evaluation team (whether it is a large multi-disciplinary team, or a few good generalists, or key specialists supported by technical advisers), all team members should get together to discuss overlapping issues and conclusions. Provision for such meetings must be included in initial budget estimates.

It is important for the team to be led by someone familiar with the issues and agencies concerned, and who is responsible for writing up

the evaluation (ie. an experienced evaluator), rather than a subject specialist. A major Somalia evaluation ran into difficulties because technical experts could not even write up their sectoral reports, and additional resources were needed in order to complete the task. An alternative option to using the team leader as content editor is to involve a professional content editor early in the process, if team members are happy for this role to be undertaken by a professional editor rather than the team leader. At all events, the writing of the final report, its structuring and language, is pivotal to the whole credibility, competence and independence of the evaluation. Has any team leader ever been heard to say that adequate time and support was set aside for this task?

Section 4

What Should Come Out of the Evaluation?

Those evaluating emergency humanitarian assistance ask, 'what happened?' and 'why?' Thus one of the objectives in writing an evaluation report is to give readers an emphatic sense of the crisis situation and its specificities, an understanding of why things happened, what worked and what did not work. Those reading an evaluation report should come away with a sense of what delivering humanitarian aid is all about. But is it sufficient for an evaluation report simply to present a coherent "big picture" and identify lessons learnt? How far should evaluators go in providing specific, implementable recommendations? Is negotiating follow-up action with responsible agencies more important than making specific recommendations?

Of course, much depends also on the nature, scope, mode and means of the evaluation concerned. Is it more oriented to audit and accountability than lessons to be learnt for next time, or *vice versa*, or both? Is it more about inputs and outputs, or outcomes and impacts, or again, both? Is it a mid- or end-term report, an inquiry made long after the event by insiders, outsiders, or a mixed team? Whatever its scope or nature, an evaluation report will maximise its potential impact if it presents its findings, conclusions, recommendations and follow-up sections separately. If readers disagree with the recommendations (or find themselves unable to implement them because of political constraints), they may be able to agree with the findings or conclusions.

Discussing findings and conclusions

It has already been noted that one of the strengths of a multi-disciplinary team is the differing perspectives it can bring to bear on issues. All team members should, therefore, be involved in discussing the findings and linking these to conclusions. Tensions which may arise between the team leader and individual subject

specialists on the nature of the conclusions can be managed more productively if the whole team is brought together to discuss findings and conclusions.

Provision (in terms of time and financial resources to bring team members together) should also be made for a further workshop to discuss comments on the draft report and any new information, prior to completing the final report. The team responsible for Study III of the Joint Evaluation of Emergency Assistance to Rwanda received around 100 pages of comments on their draft report from members of the steering committee. This detailed level of response is unlikely to be unusual, particularly in the case of large, system-wide evaluations.

> **[D]espite every effort to obtain all relevant information, there will be information which the team did not find, think to find or gain access to. There will inevitably be information that has been misinterpreted by the evaluators. The process of distributing the draft reports for comments and discussion is an integral part of evaluation (Dabelstein, 1996:291).**

It is therefore important to allow sufficient time between the draft report and the final report for team members to digest new information and feedback on the draft report.

Comprehensive discussions within a steering committee, or management group, on the draft report may not only 'smoke out' new information, including what may have been deliberately withheld from evaluators, but also contribute to a sense of ownership of the report on the part of the management group. In the case of the joint Rwanda evaluation, not all steering committee members accepted all the findings or recommendations, but all agreed that the evaluation was thorough and comprehensive (Dabelstein, 1996:291). Having confidence in the professionalism of the evaluation team and their work is important if the integrity of the evaluation report is to be preserved. For example, in the case of the Rwanda evaluation:

> **...considerable pressure was exerted by the UN Secretariat, UNHCR and the Government of France to have parts of the report removed or edited. The Management Group resisted this pressure by expressing full confidence in the quality of the work and continued to ensure the teams' independence (Dabelstein, 1996:291).**

Whether full evaluation reports should be released after completion, or only summaries, is another issue to explore. Some agencies have an express policy of not releasing full reports. From others it may be difficult to obtain reports at all, until it is too late to make any difference.

Making Recommendations

While conflict may arise over the conclusions which can reasonably be drawn from the findings, in any large-scale evaluation, conflict over the nature of the recommendations is probably inevitable. In the case of the Rwanda evaluation, there was tension between the desire to have specific, targeted recommendations and those who favoured a more discursive and explanatory approach not focused on recommendations. This tension reflects the general debate among evaluation specialists over the form in which recommendations should be made. Some would argue that evaluation reports should contain specific, implementable recommendations detailing the actions agencies should take in order to improve future performance. Such recommendations should also spell out who is responsible for implementing each recommendation, and who is responsible for monitoring whether this action takes place.

Others would urge caution, favouring findings and conclusions over specific recommendations, so as not to burden policy-makers with recommendations that could lead to unimplementable policies. If recommendations are required, an evaluation team could provide policy-makers with options rather than a single recommendation, together with an analysis of expected consequences.

Others would go further, arguing that an evaluation does not have to end with recommendations. As one member of the Rwanda evaluation wrote:

> **...I myself considered the recommendations to be the least important part of the study; what was the most important part was the analysis of the case and the issues (Adelman, 1996:2).**

Different issues may require different responses. Technical issues may lend themselves to specific recommendations in the final report. In dealing with broader issues (such as, 'why did the genocide take place?') it may be more productive to deliver the analysis to a workshop of decision-makers and evaluators which negotiates follow-up action.

Donor agencies may feel more comfortable if an evaluation focuses on findings and conclusions, rather than specific recommendations, being aware of the political constraints precluding implementation. On the other hand, donor agencies may be looking to an external evaluator to raise issues and make recommendations with the intention of stirring up debate, which those inside the organisation are unable to do.

Regardless of whether the recommendations are negotiated or independent, specific or general, if evaluation reports are to do more than gather dust on the shelves, they must be timely. In spite of the complexity of the issues involved, you cannot take years to evaluate humanitarian assistance programs because consciousness of the need for emergency assistance fades quickly. Organisations are easily able to ignore evaluation reports which are published two or three years after the events to which they refer.

It is believed that evaluations of complex emergencies are best conducted relatively quickly after the event. How then do they deal with issues of psycho-social trauma? Such issues are part of complex emergencies, yet only emerge many years (or decades) after the event. In the case of such trauma, time does not heal. Special provision for evaluation of such issues will have to be made outside regular types of follow-up, where it cannot be made within.

Follow-up action

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

But what happens next? Large system-wide evaluations raise issues which relate to a diverse range of organisations and compliance cannot be compelled. Monitoring of follow-up action is therefore important, because it provides for a level of accountability, which is otherwise missing. A well-resourced and well-structured monitoring process can strongly influence agencies to account for their response (or lack of it) to the evaluation report. Some experience in the follow-up to evaluations is noted below.

The 1996 UNWFP Liberia sub-region emergency aid policy evaluation proposed, as one part of its follow-up, a regional

workshop in West Africa with the principal purpose of refining substantive findings, conclusions and recommendations.

The 1994 SIDA Horn of Africa humanitarian aid evaluation similarly enjoyed follow-up action (Wood, 1996). Workshops were organised with the function of further examination of substantive findings, as well as refinement of the evaluative process.

The Joint Evaluation Follow-up Monitoring and Facilitation Network (JEFF), established after the Rwanda evaluation, could provide a model for institutionalising post-evaluation monitoring. However, some would argue that JEFF largely monitors rhetorical responses and would prefer to see a system which measures partial and negative learning, as well as tracking how many times a certain recommendation has been made and what has been done about it. This type of detailed impact monitoring places an additional discipline on the evaluation team, in that the evaluation report has to be evaluable. There has to be a clear link between the recommendations and the evidence in the body of the report, which supports the recommendations.

This report has collated some of the ideas and discussions which arose during the course of the Canberra meeting. It is hoped that they will make a contribution to the ongoing process of improving the evaluation of complex emergencies.

References

Adelman, H. 1996, 'Evaluation the Impact of an Evaluation: Early Warning and Conflict Management - the Great Lakes Region of Africa', paper prepared for the Joint Evaluation Follow-Up Monitoring and Facilitation Network.

Adelman, H. 1998, 'Evaluating the Evaluation of an Evaluation: The Case of the Initiative in Zaire', paper prepared for the workshop on Emergency Humanitarian Aid Evaluation, 13 - 16 March, Canberra.

Alexander, D 1997, 'The Study of Natural Disasters, 1977-1997' *Disasters,* Vol 21, No 4, Dec 1997, pp 284-304.

Apthorpe, R 1998, 'Confessions: Some actual practice in aid evaluation, some personal experience of the art', Five Houses International (Sydney, mimeo).

Borton, J 1998, 'The state of the international humanitarian system', Briefing Paper 1998(1), ODI (London).

Borton, J 1994 'The changing roles of NGOs in the provision of relief and rehabilitation assistance: northern Ethiopia and Eritrea Working paper 76, ODI (London).

Bradley, A.P 1998, 'Cost effectiveness analysis: an assessment of its application in evaluating humanitarian assistance' NCDS, mimeo.

Dabelstein, N. 1996, 'Evaluating the International Humanitarian System', *Disasters* vol. 20, no. 4, pp. 288-294.

Donini, A., Dudley, E. and Ockwel R. 1996 *Afghanistan: coordination in a fragmented state: a lessons learned report*, UNDHA (New York).

Gasper, D 1998 'Notes on the logical framework approach and the planning and evaluation of (complex) emergency relief', ISS mimeo (The Hague).

Hallam, A 1996, 'Cost-effectiveness analysis: a useful tool for the evaluation of relief operations?' ODI (London)

Kent, R.C. 1996 'The battle for the soul of aid' *Worldaid 1996: a special supplement of Crosslines Global Report.*

Kirkby, J. 1998, 'Beneficiaries and Humanitarian Assistance in Complex Emergencies' (draft), mimeo.

Kirkby, J., Kliest, T. & O'Keefe, P. 1997, 'The Complex Emergency in Somalia: UNHCR's Cross Border Cross Mandate Operation.

Lindahl, C 1998 'The management of disaster relief evaluations: Lessons from a Sida evaluation of the complex emergency in Cambodia : some excerpts' Management Perspectives International mimeo (Sweden)

Martin, L 1998 'Planning and organizing useful UNHCR evaluations: some excerpts', UNHCR (Geneva) mimeo.

O'Keefe, P. & Kirkby, J. & Convery, J & Howell, D. 1997, 'Quality in Humanitarian Assistance', ETC (UK), mimeo.

Reinharz, S. 1992, *Feminist Methods in Social Research*, Oxford University Press, Oxford.

Undro: UNEP, 1986 ' Disaster: some social aspects and their programme implications', UN (New York).

Wood, A 1996, with G. Thunstrom and Tegegne Gebre Egziabher, 'Post-evaluation workshops: some guiding principles. An Assessment of lessons from the 1995 Addis Ababa workshop on the evaluation study of Sida-funded relief projects in the Horn of Africa', Dept of Geography and Environmental Sciences, Huddersfield University (UK) (see also his 1995 learning lessons from the 1994 'Evaluation of Sida-funded relief activities in the Horn of Africa: From ideas to practice', Sida, (Stockholm).

Annex 1

List of Participants

Howard Adelman
Professor of Philosophy
Centre for International and Security Studies
Centre for Applied Ethics
York University
Canada

Raymond Apthorpe
Visiting Professor
National Centre for Development Studies
Australian National University
Australia

Claes Bennedich
Senior Evaluation Officer
Swedish International Development Cooperation Agency

John Borton
Research Fellow
Humanitarian Policy Group
Overseas Development Institute
United Kingdom

Bernard Broughton
Director
Project Design & Management Pty Ltd
Australia

Niels Dabelstein
Head, Evaluation Secretariat
Danish Ministry of Foreign Affairs

Abraham De Koch
Evaluation Officer
Office of Evaluation
World Food Program

Des Gasper
Senior Lecturer
Public Policy and Administration
Institute of Social Studies
The Netherlands

Alistair Hallam
Overseas Development Institute
United Kingdom

John Kirkby
Consultants for Development Programs
ETC (UK)
United Kingdom

Ted Kliest
Policy and Operations Evaluation Department
Ministry of Foreign Affairs
The Netherlands

Pierson Ntata
Lecturer in Sociology
University of Malawi
Malawi

Phil O'Keefe
Consultants for Development Programs
ETC (UK)
United Kingdom

Fiona Terry
Peace Research Centre
Research School of Pacific and Asian Studies
Australian National University
Australia

Annex 2
Workshop Agenda

Friday 13 March

9.15am - 10.15am	**Introductions and discussion of workshop objectives**
10.15am - 10.30am	BREAK
10.30am - 12.30pm	**Discussion of Rwanda evaluation**
12.30pm - 1.30pm	LUNCH
1.30pm - 3.00pm	**Discussion of draft DAC best practice guidelines**
3.30pm - 4.00pm	Reception at AusAID
4.00pm - 6.00pm	**Discussion of papers**

Saturday 14 March

9.00am - 10.15am	**Discussion of Somalia evaluation**
10.15am - 10.30am	BREAK
10.30am - 12.00pm	**What is a complex emergency?** Implications for the evaluation of humanitarian assistance activities
12.00pm - 1.00pm	LUNCH
1.00pm - 2.15pm	**The evaluation team** Who selects the team? What criteria are used? Processes of selection
2.15pm - 3.45pm	**Planning the evaluation** Selection of activities to be evaluated Preparing for the evaluation Pre-studies - before or after team selection? Self-evaluations Beneficiary surveys?
3.45pm - 4.00pm	BREAK

4.00pm - 5.30pm	**Terms of Reference** Negotiated or Set? General or Specific? Pre and Main Contracts?

Sunday 15 March

9.00am - 11.00am	**Methodology** Building a history of the emergency and the response Sub-criteria for evaluation Impact versus Process? Beneficiary Involvement: How?
11.00am - 11.15am	BREAK
11.15am - 12.30pm	**Writing the report & preparing the recommendations**
12.30pm - 1.30pm	LUNCH
1.30pm - 2.30pm	**Follow-up to the evaluation**
2.30pm - 2.45pm	BREAK
2.45pm - 5.00pm	**Review** Outputs Areas not covered adequately Future action

Monday 16 March*

10.00am - 11.30am	**Logical Frameworks**
11.30am - 12.30pm	**Cost-effectiveness**
12.30pm - 1.30pm	LUNCH
1.30pm - 2.45pm	**Psycho-social trauma**
2.45pm - 3.00pm	BREAK
3.00pm - 4.30pm	**Consulting victims/beneficiaries**

* This day's sessions were held with the NCDS Short professional course/ master's module participants of the 'Complex emergencies: Evaluating humanitarian aid,' namely: Ted Alan (Papua New Guinea) Andrew Bradly (Australia), James Chiusiwa (Malawi), Philip Fradd (Australia), Peta Fussel (Australia) Nomita Halder (Bangladesh), Mathias Bazi Kabunduguru (Tanzania), Oswald Maeunda (Tanzania), Heather McLean (Australia), Chola Mbula (Zambia), Sophie Pinwill (Australia),

Jane Rudland (Australia), Nattinee Marakanond (Thailand), Monhathai Rundrikabho (Thailand), Justine Severin (Australia), Kwalimu Virobi (Papua New Guinea), and Jose Viloria (Philippines).