ALNAP

LEARNING · ACCOUNTABILITY · PERFORMANCE
IN HUMANITARIAN ACTION

There will be a time when the ALNAP annual review will be able to reflect deep and enduring progress made by organisations in the humanitarian sector. This is not yet the case.

Humanitarian organisations, according to this sixth ALNAP review, still ignore lessons that should be learned from a mounting number of post-operation evaluations. They remain self-referential, striving for standards and norms that do not necessarily reflect the interests and expectations of beneficiaries; and they continue to focus on needs rather than vulnerability. These persistent failings appear to be ever more encumbered by the politicisation of humanitarian aid: aid used as "peace dividends" that in turn has become part and parcel of the global security agenda.

The ALNAP annual review explores these continuing weaknesses, with clarity, originality and insight. Even more importantly, the contributors to this year's edition proffer perspectives and recommendations that humanitarian organisations ignore at their peril. At the same time, this most recent ALNAP edition ventures into new perspectives about the humanitarian sector. Going beyond the formal system that has become quite familiar to those in the more developed world, the review rightly points to a much wider structure of humanitarianism that is the consequence of emerging powers such as India and China, and a greater understanding about the vitally important role of that vast and largely invisible system which includes host families, villagers, friends and employers who help people in cash, kind or labour at critical moments of survival.

This year's ALNAP review is required reading for those who want to come to grips with some of the central weaknesses and spectrum of solutions for the humanitarian sector as we now know it. Even more, it takes a brave step towards providing insights into a future humanitarian community that all concerned with preventing, preparing for and responding to crises of the future will have to understand.

Randolph Kent, PhD
Director, Humanitarian Futures Programme
Kings College, London

odi

ALNAP REVIEW OF HUMANITARIAN ACTION: Evaluation utilisation

ALNAP REVIEW OF HUMANITARIAN ACTION
Evaluation utilisation

Active Learning Network for Accountability and Performance in Humanitarian Action

ALNAP is an international inter-agency forum that works to improve learning, accountability and performance across the humanitarian sector. It does this by sharing lessons, identifying common problems and where, appropriate, building consensus on approaches.

Established in 1997, ALNAP has 57 Full Members and about 750 Observers encompassing Bilateral and Multilateral donors; UN Agencies and Departments; NGOs and NGO umbrella organisations; the International Red Cross and Red Crescent movement; independent consultants, academics and research institutes. To encourage self-criticism and learning, ALNAP ensures that no one type of organisation is able to dominate discussions or set the agenda. ALNAP is governed by a Steering Committee of 8 Full Member representatives selected to maintain a balance of representation between organisational types, and is serviced by a small Secretariat hosted by the Overseas Development Institute in London. ALNAP's rolling workplan of activities is developed by the Steering Committee in collaboration with the Full Members. Workplan activities are funded by subscriptions from Full Members.

www.alnap.org

ALNAP Full Members

Africa Humanitarian Action • All India Disaster Mitigation Institute • AusAID • John Borton •
British Red Cross • CARE International • Caritas Internationalis/CAFOD • Christian Aid • Ian Christoplos •
CIDA • CRS • DANIDA • DARA • DFID • DGDC Belgium • DRC • ECHO • ETC UK • FAO • GICHD •
Groupe URD • HAP International • ICRC • ICVA • IFRC • IRC • Irish Aid • JICA • MFA France •
MFA Germany • MFA Netherlands • MSF Holland • Norad • NRC • OCHA • ODI • OHCHR • OXFAM GB •
People in Aid • ProVention Consortium • RedR • SC UK • SC US • SCHR • SDC • Sida • Sphere •
John Telford • Tufts University • UNDP • UNHCR • UNICEF • USAID/OFDA • VOICE • WFP • WHO •
World Vision

Global welfare: A realistic expectation for the international humanitarian system?
Hugo Slim

Proportion and distortion in humanitarian assistance
Tony Vaux

The utilisation of evaluations
Peta Sandison

ALNAP REVIEW OF HUMANITARIAN ACTION
Evaluation utilisation

The opinions expressed in this publication are those of the authors and do not necessarily represent the views or policies of ALNAP or any of its members.

The sections of this publication can be downloaded from the ALNAP website www.alnap.org

Contents

List of acronyms and abbreviations

ALNAP   Active Learning Network for Accountability and Performance in humanitarian action

CAFOD   Catholic Agency for Overseas Development

CAP   Consolidated Appeals Process (UN)

CfR   Cash for Relief

CHAP   Consolidated Humanitarian Appeals Process (UN)

DEC   Disasters Emergency Committee

DFID   Department for International Development (UK)

DRC   Democratic Republic of the Congo

ECHO   European Community Humanitarian Office

ERC/USG   Emergency Relief Coordinator and Under-Secretary-General for Humanitarian Affairs (UN OCHA)

FAO   Food and Agriculture Organisation

GHD   Good Humanitarian Donorship (initiative)

HAP   Humanitarian Accountability Partnership

HIC   Humanitarian Information Centre (UN)

IASC   Interagency Standing Committee

ICVA   International Council of Voluntary Agencies

IDP   internally displaced person

IFRC   International Federation of the Red Cross and Red Crescent

INGO   international non-government organisation

JEEAR   Joint Evaluation of Emergency Assistance to Rwanda

M&E   monitoring and evaluation

MSF(H)   Medecins sans Frontieres – Holland

NGO   non-governmental organisation

OCHA   Office for the Coordination of Humanitarian Assistance (UN)

OFDA   Office of US Foreign Disaster Assistance (USAID)

RBA   rights-based approach

RTE   real-time evaluation

TEC   Tsunami Evaluation Coalition

ToR   terms of reference

UNHCR   United Nations High Commission for Refugees

UNICEF   United Nations Children's Fund

USAID   US Agency for International Development

WFP   World Food Programme (UN)

WHO   World Health Organisation (UN)

Preface

Last year's *Review of Humanitarian Action* (RHA) alluded to some positive developments with regard to evaluations of humanitarian assistance. We now know that evaluations are more engrained in the system than before; there are more innovative evaluative approaches including Real Time Evaluations (RTEs); the system has demonstrated its ability to conduct system-wide Joint Evaluations such as the TEC; and there has been a small improvement in evaluation quality. All of this is laudable but there is still a big question mark over whether these improvements have been translated into improved performance. Indeed, the findings from the Tsunami Evaluation Coalition clearly indicate that there is still a long way to go.

I think it is therefore very appropriate that this sixth edition of the RHA directly addresses why the lessons from humanitarian experience have not been learned and therefore have not acted as catalysts for improving humanitarian action. Chapter 3 gives a well-argued explanation of the myriad of reasons why clear and common sense recommendations from evaluations are not acted upon. While it makes for sober reading, I believe it is crucially important to understand the kind of barriers that exist, so that we can develop approaches and strategies that take account of these problems. Indeed, the chapter outlines a number of practical steps that are relevant for evaluators and evaluation managers alike. This kind of practical advice has the potential to make a real difference in take up and utilisation of recommendations.

Chapter 2's evaluation synthesis provides an insightful account of the distortions that occur in the field as the result of a humanitarian system that still does not function fully in accordance with humanitarian principles. The evaluations support the proposition that the global allocation of aid is not based on a common understanding of 'need' and provide clear evidence that the humanitarian system is therefore not properly directed. Again, the chapter provides some very clear examples of what actually happens and presents some clear messages which will help guide both practitioners and policy-makers in their work.

I am sure that many of us are so close to the issues and challenges presented in the above 2 chapters that we sometimes find it hard to see the wood for the trees. That is

why this year we have asked Hugo Slim to stand back on our behalf and present us with a bigger view in Chapter 1. What can humanitarians realistically hope to achieve in a global humanitarian system that is so inextricably tied up with international politics? Is a global welfare system a realistic aim and what should our attitude and relationship be to other powerful informal humanitarian systems that exist in the world today? This chapter provides us with an illuminating and unusual analysis that I am sure you will find exciting and challenging.

I would also like to thank all those ALNAP Members who provided such valuable advice and comments to the authors of this edition; to those agencies who made case study material available; and to those who provided evaluative reports for this edition. As usual, the willingness of ALNAP Members to give their time and share their experiences demonstrates the vitality and the goodwill in the network and ensures that the RHA is well grounded and relevant.

Eleanor Monbiot

Chair of ALNAP

Global welfare: A realistic expectation for the international humanitarian system?

Hugo Slim

1.1   Introduction

Expectations of international humanitarian aid have changed dramatically in the last twenty years. Aid budgets and aid organisations have got much bigger. Humanitarian action has come to occupy a more central place in world politics, both in the policy of powerful states and in increasingly powerful transnational movements in global society. The practice of humanitarian action – its values and its technique – is seeking to become more professionalised. The growing adoption of a political theory of rights, rather than charity, to justify humanitarian action has newly politicised humanitarian help in the minds of givers and receivers alike, both of whom increasingly expect aid to be given as an obligation under international law. In many ways, this growth can be interpreted as the emergence of a nascent global welfare system around war and disaster.

If there has been such a dramatic development in the humanitarian system in the last twenty years, what can we expect of the next twenty years? Jan Egeland, the United Nations Emergency Relief Coordinator, is clear about what he expects: 'We need a humanitarian system that is able to respond reliably, effectively and efficiently across the full range of emergencies. Now more than ever, humanitarian aid must be the responsibility of all nations for the benefit of all nations' (Egeland, 2006). Egeland's expectation is nothing less than global welfare and many other humanitarians share his view. Humanitarian agencies want to get bigger and better because they want to reach everyone in need and reach them well. Behind many of the efforts to expand and improve the humanitarian system is a desire to see some form of high-quality humanitarian support for everyone affected by war and disaster. Whether people are suffering as a result of mudslides in Central America or vicious displacement in Darfur, the humanitarian system wants to help them. Global wealth suggests that it can and global morality says that it should.

These are ambitious expectations for the humanitarian system and are not without real political, technical and organisational obstacles. Meeting them will inevitably require a focus on two areas – politics and practice. Any overall improvements in the humanitarian system are dependent, above all, on changes in the realm of politics where humanitarian aspirations can be enabled, obstructed or distorted. The realm of humanitarian-agency practice in which humanitarian organisations can professionalise, innovate, coordinate or fail is the second area that determines

improvement. Together, the outer realm of politics and the inner realm of practice are the separate and overlapping fields of humanitarian action.

ALNAP has now been reviewing humanitarian progress for the last five years. As ALNAP and others begin to assess the performance of the system in the next five years, it is surely sensible to step back and ask how realistic such high expectations are within international society or, at least, how we should calibrate our expectations as we move ahead to meet them. Perhaps ALNAP members are right to be judging the system against the highest expectations and the fullest vision of success. Their goals should certainly be informed by an ideal of some kind. Maybe ALNAP should temper such long-term ambition with more modest interim dials and benchmarks on their evaluator's dashboard. Or, perhaps the differences within international society are such that it is unwise ever to expect a functioning global system for anything that involves such massive competing interests as war, disasters and the world's fast-emerging environmental crisis.

In this chapter, I want to complement the detailed analysis of the system which follows in Chapters 2 and 3 with a more general examination of the political and practical assumptions which lie behind the call for global welfare in order to see if this expectation is realistic or misplaced. To do this, I will take a very broad view of the humanitarian system as it currently exists and performs in international society. From this wider view, I will suggest that humanitarian progress is best served by a deliberately strategic approach to very specific mid-range improvements which we can reasonably expect.

The rest of this chapter isdivided into three sections.

- Section 1.2 looks at a range of different possible expectations of the humanitarian system. It discusses various assumptions about the prospects of global humanitarian progress, and shows how each one inevitably leads to different measures of success depending on differing views of what is politically and practically possible.

- Section 1.3 looks at the system we have now. It gives a general description of the formal international humanitarian system, and its current performance problems, and then also describes other less well-recognised but very powerful informal systems of humanitarian action. These informal systems complement

the official system in many people's lives, but do not actively collaborate with it and may have interests in undermining it.

- Finally, Section 1.4 suggests taking a five-year view of what we can reasonably expect as humanitarian progress, and makes some general recommendations about how the system might usefully set, monitor and approach ten keynote objectives for improvement.

But, first, it is important to examine how much improvement in humanitarian action we can reasonably anticipate in the world around us.

1.2 Possible expectations of the humanitarian system

1.2.1 Humanitarian progress

There has long been a discussion in international relations about what exactly we can expect from international society. Is it realistic to judge international society on the same terms as we would a nation-state? Can the world ever be as organised and regulated as a single country? Can common agreements, a single rule of law and an effective form of government be upheld across the globe?

The whole impetus of the United Nations movement and much international NGO work is based upon a belief that such things are possible and that we are rightly, if falteringly, evolving towards a morally responsible form of global order. Many humanitarians naturally feel that, as a key part of this global political evolution, we should expect to see real progress in the fair and efficient provision of humanitarian aid worldwide. But should we? Is it realistic or helpful to have such expectations?

In its current form, the belief in reasonable progress towards a just global order is most obviously a product of the European Enlightenment. Although, of course, it has always had equivalent religious manifestations in some Christian and Islamic hopes for a world order overseen by Church or Caliph. In the Enlightenment and

throughout the modern era, many people have believed in the possibility of just global government and orderly relations between states.

The seventeenth-century Dutch scholar, Hugo Grotius, wrote his *Laws of War and Peace* in 1625. This treatise is widely regarded as the modern starting point for international laws and treaties that have increasingly bound nations into a reciprocal form of mutual respect and obligation. In 1795, the German philosopher Emmanuel Kant wrote his famous essay, *To Perpetual Peace*, which supported the conviction that nations would one day be able to live in an orderly law-abiding world based on common interests and shared values. More recently, in 1919, US President Wilson and others set up the League of Nations in Geneva, with the intent of making peace, prosperity and global order a reality – a vision that was later taken over by the United Nations. Contemporary philosophers like Richard Rorty and Amartya Sen have focused on the ideas of solidarity and rights as the political cement with which to build the necessary global consciousness to develop just international relations.

But these ideals of an orderly and well-governed world have always been challenged by other more realist or pessimistic visions. Many people think it foolish to expect that the world will ever become the global equivalent of one big nation-state with a responsible and effective global government of some kind that is accountable to all its citizens. Instead, they believe that international society will always be the ultimate arena for competition much more than for cooperation. Others think that the most we can expect from international society is a sort of benign international anarchy where states and peoples can make and keep some rules but never really be consistent, well organised or beholden to a single global authority. They might also add that such anarchy is the best way to defend against the possibility of a global government that could become autocratic, unrepresentative and hard to change.

A discussion of international relations is important to any discussion of the performance of the world's humanitarian agencies because international politics is implicit in almost every page of the chapters that follow. How we think the world works, or could work, will govern what we can reasonably expect of global humanitarian action. What we believe about the realistic prospects for global order, international consensus, equitable resource distribution and recognised legal authority determines how we will evaluate the current international system of humanitarian action. In short, how much we think the humanitarian system can be changed and improved depends on what we believe is possible of international society.

1.2.2 Political expectations

Depending on our view of international society, there are three main kinds of political expectations we might have of humanitarian action within the global system. Each of these perspectives is likely to make us judge the current humanitarian system rather differently.

1 An optimistic political view of international society regards the world as now willing and able to deliver impartial and effective humanitarian aid. This view leads to a harsh judgement of current humanitarian action because, as this review and all its predecessors show, it is obviously neither fair nor efficient.

2 A cautiously progressive view of international society sees it as gradually getting closer to a just global order. Anyone holding this view will probably be heartened by the steady if imperfect progress that humanitarian ideals are making in global consciousness, and the practical effect of humanitarian action in the lives of an increasing number of people. All the ALNAP reviews to date have identified serious problems in the politics and practice of humanitarian action but they also talk of progress in several areas.

3 A pessimistic view regards a just global order as unlikely. In this case, we may evaluate current humanitarian action rather favourably. We may be amazed that some governments are prioritising it at all, albeit sometimes distorting its ideals as they do so, and that so many agencies are trying to establish humanitarian work as an essential international obligation and a legitimate profession. We would not expect dramatic results from any project working in these conditions but might praise the almost foolhardy spirit of humanitarians to keep on working and trying to improve.

Reading between the lines of ALNAP's five reviews so far, and listening to statements of humanitarian leaders like Jan Egeland, it is the first of these views which tends to predominate in the humanitarian system's criticism of itself. Such optimists clearly believe that conditions are now riper than ever for achieving dramatic progress in a global humanitarian system. For the optimists, the unprecedented sufficiency of resources, communications, know-how and moral consensus make this a unique moment for realising the ideal of global welfare in war and disasters. Only the stubborn ox of political will seems to be evading the task.

However, with victories on landmines and the International Criminal Court already in their bag, humanitarian optimists think that even this most obstinate beast can be harnessed with sufficient public pressure and the realisation that disasters will be more frequent and much bigger as the environment wears out.

People who take this optimistic view and think that global order and shared international interests are now more realisable than ever must share Tony Vaux's obvious frustration in Chapter 2 and judge the recent performance of the international humanitarian system harshly. Vaux's excellent overview of evaluations from the last five years makes clear that current international and national politics mean that the system continually fails to work in an orderly, ethical and efficient fashion. Instead it is regularly distorted in the allocation of its resources, obstructed in its access or unprofessional in its practice.

Those who take a more cautious but still positive approach to the evolution of international society must expect any progress towards sufficient, fair and high-quality humanitarian action around the world to be slow but would still expect the trend in performance to be essentially upwards. From this perspective, the overview in Chapter 2 is not too depressing. Despite perennial political distortions and continuing operational failings, the humanitarian system seems to be in great demand by certain governments and publics who want to fund it and use it more than ever to help millions of strangers around the world.

Less clear is whether the people on the receiving end of the system are getting what they want and need. Many are not getting nearly enough while others are not being asked if what they did receive was useful. However, there is no doubt from the various evaluations that have been examined in this ALNAP review, and from the general discussion around humanitarian action today, that the profession is more self-conscious and sophisticated in its understanding of itself than ever before. It knows what it needs to know and do to improve. The problem is – as both Tony Vaux and Peta Sandison point out in the following chapters – that the system is not very good at acting on what it knows. Good intentions get crowded out and make learning and improvement genuinely difficult locally and globally.

From a pessimistic perspective of the prospects for an orderly international society, things look different again. If we expect the world always to be essentially anarchical and dominated by a diverse mix of competing powers, interests and ideals, then we

would draw a more tolerant conclusion about the continuing political distortions and operational failures revealed in this new ALNAP review. We would not judge the system and its agencies so harshly. We would have to conclude that the very idea of organised, impartial and effective international humanitarian action is a very high ideal never likely to be fulfilled as we would wish. As a result, we might observe that international humanitarian action is neither all bad nor all good but that it is, at least, the preoccupation and preserve of a few powerful Western governments and their publics who believe in its ideals and pay most of its bills even if they also use it as a part of their wider foreign policies.

The inevitable range of competing and diverse political interests in the world means that we should not be surprised at the number of other powerful states who ignore this particular international humanitarian project. Many do not join in with it, are excluded from its Western club or actively seek to block and undermine it. Most obviously, states like China, Russia, India, Brazil and several rich Muslim states are notable by their absence from the centre of the formal system and do not pay much towards it. In their own systems, they too are highly selective about the wars and disasters with which they actively engage. Fairness and efficiency are not key criteria for them either.

India and Pakistan limit their humanitarian engagement but make significant peacekeeping contributions to the UN and play important roles in discussions of natural disasters from which they often suffer. Arab and other mainly Muslim states tend to focus their humanitarian advocacy and donations on wars which involve Israel or the USA fighting Muslim populations. They keep relatively quiet about and give very little to wars like those in Darfur and the DRC. China and Russia both explicitly equate much international humanitarian action with Western imperialism and prefer to focus on their own bilateral trade and aid relations to secure their own growing interests in Africa and Asia.

In short, we cannot evaluate progress in the fairness and effectiveness of international humanitarian action unless we have a baseline of what we can reasonably expect from it, within the wider political context. It would be the same with the speed of a 5000-metre runner. It would be impossible to judge her performance without knowing if she were running uphill all the way, over stony ground or along a smooth, flat track. A speed which would be slow on the flat might be pretty good if running uphill most of the way. Similarly, before we judge the humanitarian system we must be honest about our own political assumptions about

international society. Is the humanitarian system being asked to perform uphill, on stony ground or on a nice, smooth track? How we answer this question will determine how we judge the performance of the system.

1.2.3   Agency and recipient expectations

One of the interesting things about the last fifteen years of humanitarian action has been the determination of some parts of the system – initially the NGOs and then government donors and UN agencies – to be more explicit about what they expect from the humanitarian system through the Code of Conduct, various quality standards and the Good Humanitarian Donorship initiative. ALNAP's five reports to date have also made an important contribution to the clarification of the system's expectation of itself. The very process of doing syntheses of evaluation findings from the whole sector has helped to reveal many assumptions of perfection that exist in the industry on a whole range of issues. What evaluators keep identifying as wrong clearly reveals what they think is both right and possible.

All together, the standard setting, evaluation and discussion of improvement within the humanitarian community leaves an unambiguous message: that the global expectations of the humanitarian profession are very high across a vast range of people's needs and that they are, in many ways, equal to a standard of service expected within a modern state. Deep down, it seems that many people inside the humanitarian community expect the international humanitarian system to be a global emergency service that is as fair and effective as the combined ambulance, police and fire services of a modern state. This is a deliberate exaggeration but it serves to reveal the high standards expected by many humanitarian advocates as a simple matter of equal rights for all people worldwide.

Equally interesting, however, and a recurrent theme in the pages that follow, is that we really have a view of only the expectations of the providers in this system. Still, after many years of emphasising their importance, we have only the slightest inkling of the expectations of the recipients of humanitarian aid and protection – the people who actually suffer from war and disaster. These people are the real customers or end-users of humanitarian action. Any business or profession knows that it cannot hope to succeed without knowing what its clients, patients or customers actually need and want. If the last few years have been critical for revealing what

humanitarians expect from the system, their expectations must now be compared rigorously and routinely with those on the receiving end.

This is where the current discussion of need – which is rightly given so much importance in Chapter 2 – is critical. The international humanitarian system needs some credible measure of global needs which gives a continuous sense of its targets and its bottom line. Other parts of the international system – like the World Bank and UNDP – are able to estimate and aggregate a variety of socio-economic standards like the cost of living, parity of purchasing power between nations, the quality-of-life indicators and the human development index. It should not be difficult for humanitarians to work with these institutions and regional organisations to develop a similar scale for calculating generic need in war and disasters, then costing them according to different settings.

Such a system would be more sophisticated and more transparent than the current UN Consolidated Appeals Process (CAP) which is so subjective and potentially self-interested. The CAP system involves those institutions with most interest in receiving the funds also being most responsible for calculating the needs. A pre-designed empirical calculation of need would be more scientific. It would allow the system to anticipate and prepare more coherently in advance of crises. It would also counter skewed public sentiment with scientific precision to prevent the absurdity of such imbalances as those between the Tsunami response and the DRC response. A humanitarian system that has aspirations to be a global system must have a clear and convincing way of describing its priorities, even if they are not met.

1.3   Global humanitarian systems

If the previous section describes the different expectations we may have for global humanitarian action in international society, it is now important to compare these with the system we actually have at the moment. To do so, I will paint a general portrait of the formal international system as it exists in international society today, and then also make some smaller sketches of several other, more informal global aid systems which are also emerging fast around war and disaster.

By the formal system, I mean the mainly Western-funded humanitarian system which works closely within or in coordination with the international authority of the United Nations and Red Cross movements. It therefore includes the great number of non-governmental organisations who take some funds and some policy direction from governments who invest mainly in this system, as well as also many other funds from individual private donors. By the informal system, I mean other large and organised aid flows which go to people affected by war and disaster but which do not go through the formal UN-centric system.

1.3.1 The formal international humanitarian system

Stepping back from the formal humanitarian system, the first thing we see from afar is the attempt by about twenty morally earnest and politically driven European and North American states to develop a rudimentary global welfare system that can provide a high-quality safety net for people suffering from war and disaster around the world. Already, they have achieved remarkable things. Almost every war or natural disaster now receives the attention of this nascent welfare system, usually mediated in some way through the international authority of the United Nations or the Red Cross movement. Significant and routine transfers of resources, equipment and expertise mean that millions of people receive help from this system every day.

Looking back further to the roots of this endeavour, we can see its origins in a variety of earlier nineteenth-century European and North American engagements with war, colonialism, anti-colonialism and the poverty of their own societies. From its mixed genesis in the Red Cross movement, international mission, colonial administration, independence movements, social reform and radical solidarity groups, the humanitarian system is now organising itself as an emerging form of global welfare. This emerging system is coming under increasing pressure to model itself on the elaborate needs-based and means-tested state provision that has become a feature of the political traditions of these countries. In line with progressive welfare theory, the formal humanitarian system also aims to inculcate a notion of rights, duties, self-help, economic opportunity and accountability into the societies in which it works.

The formal humanitarian system has a range of operators. It is currently managed mainly by the UN (an explicitly governmental organisation and the prototype of an

international civil service) and the Red Cross movement (an international network of government auxiliaries and volunteers). It also relies on a growing number of more or less independent NGO agencies which use both private and government money. All these implementers receive firm policy instructions from the humanitarian departments of their donor governments, although much of this policy is worked out in a continuous policy dialogue between donors and providers. UN agencies are complete subcontractors of the system while non-governmental and Red Cross/Crescent organisations operate independently or as semi-independent subcontractors.

The donor governments at the core of the system use this emerging global welfare system as a moral end in itself to help people in distress but also instrumentally as one of several foreign-policy instruments to secure their own democratic political influence wherever possible. In doing so, they work hard to apply humanitarian aid and ideals to stop other people's wars and to mitigate their own. Humanitarian aid is often a first step in a much deeper political project to transform a society in crisis. To some humanitarians, this dual use of humanitarian aid is a sinister mixed motive, but to the governments themselves, many of their taxpayers and not a few aid recipients, it seems like a natural win–win: a strategy for saved lives and improved politics. UN agencies and most NGOs share similar double objectives wherever they engage.

As it has developed in recent years, the formal system has adopted many of the routine practices of modern welfare provision in Western states. It has become increasingly concerned about its principles and the quality of its services. It wants to make decisions fairly, do things well and be able to prove its results in some good scientific reporting. To this end, it has tried to adopt a modern bureaucratic model of management and organisation to shape and sharpen its fast-growing organisations whose smaller antecedents were originally based on simpler notions of charity, volunteerism, activism and social service.

This organisational transformation has been deeply challenging to many humanitarian organisations of the late twentieth century – just as it was for the many nineteenth-century educational, medical and labour charities who were absorbed, modernised or rendered redundant by the emergence of state welfare in twentieth-century Europe or its commercial equivalent in the USA. Many of today's agencies – UN and NGO – continue to resist what they see as absorption into an essentially

government system that they cannot control. They have been particularly alarmed at possible co-option by these governments into ambiguous political agendas around war and disaster. In response, all UN agencies and most NGOs have settled for a compromise whereby they 'roll with the punch' on the big political emergencies but also speak out about forgotten emergencies with equal vigour to assert their commitment to global fairness. A few – like Medecins sans Frontieres and Oxfam GB – have gone further and refused Western government funds in wars where the donors themselves have been outright belligerents.

As the system has grown, the number of agencies has grown with it. Many new national and international humanitarian agencies have been formed so that the humanitarian profession is now crowded. Commercial companies and military forces have also established their credentials as contractors for humanitarian funds. But, still, the so-called 'international humanitarian system' has remained deeply parochial in its European and North American core. These governments have shown no obvious interest in going on a recruiting drive to bring in new governments as members of the system. Counter-balancing efforts by a few progressive NGOs to form wider global movements of popular humanitarian concern and funding have not yet risen to match the resources and power of the core humanitarian states.

This quick sketch of the formal system as it looks today reveals how politically and culturally specific this system really is. The formal humanitarian system is really 'our' system – I think I can assume that most readers of the RHA will be from within this system, not from the informal systems mentioned below. Although this UN-centric system reaches out to many different cultures and works within them, it is a very Western affair, as the TEC evaluation indicates so clearly in its analysis of Western aid overpowering local institutions. It works wherever it can in international society but is not really owned by all of international society. Indeed, a wider look at international society reveals that this very European and North American system is not the only global humanitarian project. There are others which are possibly even bigger and faster but not necessarily fairer or better in terms of impartial needs assessment, professional technique or capacity-building.

1.3.2   Informal international humanitarian systems

The formal humanitarian system may well be seen by states and individuals not party to it – usually by choice more than discrimination – as a relatively closed and particularly Western system. Many of these other states and individuals obviously prefer to provide humanitarian aid through other highly organised political funding systems or looser social networks. ALNAP is not in the business of evaluating these other systems, but it is important to be aware of them to see if they offer a site of fruitful learning or collaboration for the formal system, and also to understand how they might undermine Western ambitions for a formal global system.

First, there is the Islamic system that operates in the less bureaucratic but often even more welfarist Muslim tradition of *zakat*. Muslims make up the largest proportion of humanitarian aid recipients today and Muslim states and Muslim people give billions of dollars in aid when they are politically or emotionally moved to do so. Like the Western system, the Islamic system is a mix of government and private donors and is also highly political in its giving. It may not give much to Darfur, where Muslims are killing Muslims, but it certainly gives lots to Palestine, Lebanon and Iraq if Israelis or Americans are killing Muslims. The Islamic system can give freely and fast (if not necessarily professionally), without the bureaucratic statist culture of the Western system. For example, while the Western system has spent years discussing, researching and piloting the morality and efficiency of giving cash relief in emergencies, Hizbollah gave out $12,000 in cash to every war-affected household in southern Lebanon within days of the recent ceasefire.

Direct cash giving is a feature of another important humanitarian system in international society – the widespread practice of sending remittances. Most remittances support the development needs of peacetime families, but a significant amount contribute to and are increased during the emergency relief efforts of individual families in many parts of the world when war or disaster strikes. Partial by nature, this system is not strictly humanitarian either, but it can have a significant effect in crisis alleviation. Recent research by Tufts University shows that the dramatic decrease in remittances from Sudanese migrant labourers in Libya to their families in Darfur, after border closures in 2003 and Libyan restrictions on migration, was a massive blow to the livelihoods of thousands of families (Young et al, 2005, chapter 5). If sustained, these remittances could have been a vital source of

aid. The Tsunami evaluation also identified the importance of remittances without being able to quantify the flow of funds involved (Flint and Goyder, 2006, pp 27–28).

Several states are also much more active outside the formal system than within it. China's increased engagement with Africa and Central Asia has a humanitarian dimension which is strictly bilateral but often critical in the recovery and reconstruction stage as it provides new roads, power plants, other vital infrastructure and (highly conditional) loans. In certain places, Russia plays a similar role and India seems curious to start. Regional powers like South Africa, Nigeria, Malaysia, Iran and Venezuela may also play aid or rehabilitation roles that escape the accounting process of the Western system.

Finally, of course, as the Tsunami evaluation and the recent war in Lebanon showed so clearly, there is the enormous front-line role of local humanitarian systems. This vast and largely invisible system includes: host families who take in IDPs; neighbours who rescue their fellow villagers and families; friends and employers who help people with cash, goods or labour at critical moments of survival.

Looking at the variety of international humanitarian systems in this way, it is clear that Western-funded humanitarian action occupies a well-organised and far reaching niche in international society. It makes a point of working through the United Nations, the Red Cross and formal NGOs and so has international legality and legitimacy in the eyes of a majority of states. It has a particularly modern scientific way of operating but does not live up to all of its high ideals. At the moment, it has not convinced many other powerful states and movements within international society to join it fully. Many seem to prefer to operate their own systems even if the Western system takes much of the limelight in the global media.

Many people within Western humanitarian agencies are mesmerised by 'our system' and its apparent importance, but other political forces may remain untouched by it or infuriated by its pretension. Some very serious powers have no desire to see the Western system become the global welfare system of tomorrow. This has important implications for those who have global expectations of the formal system.

1.3.3   Expectations of humanitarian practice

This description of the outer political realm of the formal system is incomplete without a similar picture of its inner organisational realm of practice. As with politics, so too with practice – the expectations of evaluators are high not least because many operational problems cannot be blamed on politics. Certainly, as Vaux points out in Chapter 2, political pressure bears down on humanitarian agencies when donors are in a hurry, as in Afghanistan or the Tsunami response. Resistant armed groups and governments regularly stop agencies from working as they want to. Too much money or too little money from donors each brings problems of its own. Yet, the five-year trail of ALNAP evaluation syntheses still consistently suggests that, regardless of politics, agencies could much improve their day-to-day practice. This is undoubtedly true.

ALNAP's findings repeatedly confirm not just the same structural political flaws in the global fairness of humanitarian action, but also perennial problems in its professionalism and efficiency. The problems of neither politics nor practice are new. The most superficial glance at humanitarian history reveals that the modern humanitarian movement has always struggled on both fronts. If one reads accounts of Save the Children's relief operation in the Russian civil war of 1921, UN work in the Ethiopian famine of 1984 or this year's joint evaluation of the Tsunami response, each one laments the political interference of donor and recipient governments and the constantly debilitating operational effects of, for example, high staff turnover alongside the other constant operational difficulties.

Evaluations of the formal humanitarian system continue to show how the combined system struggles in highly strategic areas of operational response, as well as with political constraints. Chapter 2 cites the trend of the formal system continuing to embody an unbalanced portfolio of sectoral expertise so that, for example, food is over-emphasised at the expense of other critical needs. Provision of shelter is routinely amateur and ill informed. Camp management, protection, water and sanitation are similarly underpowered. It remains a mystery why, in an apparently entrepreneurial and still unregulated system, so few innovative and dynamic new agencies have appeared to fill the many gaps and opportunities in humanitarian response. Maybe it reveals what many in the French tradition fear: that the system has passed its innovative phase and is now creaking towards bureaucracy and overfed inertia. If so, the UN's ongoing humanitarian reforms and the many other

improvement initiatives need to challenge the culture of the system as well as its organisation and technique.

Many of these gaps were also highlighted in last year's Humanitarian Response Review conducted by the UN, which found continuing uneven expertise, skewed resources, inadequate preparedness, poor inter-operability and weak coordination in the system (OCHA, 2005, especially chapter 2). Previous talk of coordination has not done much to reorganise the spirit and resources of the sector more effectively on the ground. Perhaps clusters will now prove to be a magic word that enables people to organise the system's offer more evenly and expertly. Strong leadership will be an important part of this.

Beyond technical expertise within sectors, Vaux shows in Chapter 2 that expertise across the phases of crisis – relief, recovery and development – also remains strangely immature in a profession that has recognised a role for itself in all three phases for several decades. The key area of livelihood support – which has the potential to unlock the bolts between these three phases and avoid the professional anathema of dependency – is similarly undeveloped in many agencies. Particularly notable is the strange fact that the humanitarian sector has ideologically resisted the most obvious strategic partner in livelihoods work – the private sector – and tried unhelpfully to make income and markets a sacred and protected part of humanitarian space in times of crisis. Perhaps Grameen Bank's Nobel Prize success will encourage humanitarian orthodoxy to become increasingly porous to commercial values and techniques in supporting people's recovery.

A particularly bizarre aspect of agency performance is the way in which core parts of the humanitarian NGO and UN establishment set out elaborate values and standards which they seem to cherish on paper but forget in practice. Tony Vaux's bemusement at the widespread lack of operational and evaluative reference to the Code of Conduct is understandable. Why make codes if you do not use them daily in your work? The Humanitarian Response Review also noted the high talk-up but low take-up of the SPHERE standards. These standards have secured a few high-profile champions in certain agencies but are not yet emerging as routine industry standards in any way. But they may do so in future.

Expectations of good practice must also be disappointed by the important findings from Peta Sandison's research in Chapter 3, which shows that agencies still often fail to learn from evaluations of their operations. Sandison's findings conclude that

humanitarian agencies find it hard to use evaluations to make real improvements to their operational practice or to challenge political distortions systematically. Evaluators and those who commission them are both at fault, it seems, because evaluations are seldom designed to be user-friendly. Nor do they take enough account of context to apportion accurate responsibility for humanitarian failures between agencies and political factors. Many evaluations are also prone to criticism more than encouragement, fail to include a learning process or are administered as routine accountability 'rituals'. Although agencies profess to hold evaluations as crucial to improvement, they do not use them well and there is no systematic link between evaluation findings and wider policy making and decision making in the sector.

Such confused utilisation of evaluation findings makes one wonder whether dramatic practical improvement is a realistic expectation to have of humanitarian agencies within the terms of their current political contract with donors and receivers. Optimistic internationalists must surely ask why agency performance is not yet subject to rigorous and regulated international oversight. But it seems that this is not yet politically possible or politically important either. Peta Sandison makes clear in Chapter 3 that performance is not actually a strategic issue for humanitarian agencies, when nobody is really on to them about it. In the meantime, senior managers may be easily diverted by too many competing issues to make performance top of their to-do list, or maybe they are more interested in funding, status and coverage rather than quality. Maybe they are even in the grip of some deep denial or fatalism about their operations which they feel will always be thwarted.

Why does practice continue to be such a problem? Is it because the humanitarian system is a particularly bad learner, as ALNAP studies have often suggested? Is it because all the things which agencies try to do are genuinely extremely difficult and very hard to learn at system level? Or is it because these things are particularly difficult when attempted by an essentially extraneous force like the Western humanitarian system? Perhaps the issues of relief, recovery and development are not deep problems in themselves, but difficult only if you are coming at them as outsiders.

Is our attitude to practice also part of the problem? Maybe *being there* rather than *being successful* is somehow more important to the humanitarian imagination and its ideal of solidarity. Frankly, successful practice is seldom rigorously demanded of us.

Nothing usually happens to us if we are bad at our jobs. Contract culture means that failures are left to fade away with the end of a contract and are seldom addressed sharply with discipline or dismissal. When operations go badly, NGOs can retreat to the warm rhetoric of shared struggle while the United Nations remains in a world of its own, where the grim culture of the great bureaucrat lives on. If attitude is a factor in disappointing practice, then the humanitarian community does indeed need cultural as well as organisational reform – a nettle which has not been properly grasped in the UN reform process.

Many people in the humanitarian system do care deeply about its practice. Several common problem areas (which have been previously and regularly identified in previous RHAs) are being very actively addressed in various interagency projects and initiatives. The Humanitarian Accountability Partnership (HAP) is working hard on agency accountability, particularly to those who need and receive aid. The UN reform process and its new cluster approach is a well-supported and serious effort to address perennial problems of coordination and the optimal distribution of resources and expertise. For their part, donor governments are committing themselves to incremental steps towards their definition of Good Humanitarian Donorship which, among other things, takes important account of the principle of needs-based impartiality. SPHERE continues to consolidate and develop what constitutes good technical standards of professional work in key sectors, and People in Aid is building a clear sense of good practice in the recruitment, retention and development of staff – one of the chronic weaknesses of the profession.

Yet, at the moment, there is a lack of collective vision and synergy between all these efforts. Moreover, there is no single unit charged with a responsibility to oversee and report on all the separate initiatives, the progress they may be making in the system and their impact on the ground. Still, other than through ALNAP, we have little idea of how we are getting on overall. Still, we have no common and simple measures of global need and success. Still, there is no independent or semi-independent office or organisation that has the oversight and mandate to judge, criticise and praise the system from outside. The Humanitarian Response Review (OCHA, 2005) has recommended common benchmarks, but these are mainly around operational process not outcomes, and they are still seen as a matter of self-regulation by the IASC and its agencies rather than the work of a genuinely independent regulatory authority.

1.3.4   Reasonable expectations of the humanitarian system

So what expectation is it reasonable to have of the international humanitarian system, its prospects within international society and the performance of its organisations in war and disasters? Should we expect a lot or a little from the international humanitarian system, given the persistence of its political and practical problems? It seems that the humanitarian community itself is still torn over what it can expect of humanitarian progress. Deeply pragmatic in one part of themselves, agencies fast learn from operational experience not to expect too much from international society. Yet, equally infused with more than their fair share of idealists and campaigners, many humanitarian agencies constantly demand more from international society and seem to think that global perfection is possible.

ALNAP's five studies to date certainly give out a mixed message. Practitioners and evaluators of humanitarian action are at once pessimistic and ambitious. Many seem to agree with Tony Beck and John Borton, writing in 2003, that 'many factors that determine whether or not humanitarian action is successful or not are largely outside the control of humanitarians'. However, many also seem to elide this view with one which still has significant expectations of this flawed political system. The implication in most humanitarian discourse is that international society is indeed amenable to ordering according to humanitarian values. Why else would humanitarians continue to demand this?

1.4   A five-year view

It seems to me that the humanitarian profession should take a mid-term view of its possibilities and performance and not get too enthralled to ambitious long-term hopes for international society. Another five years is not a bad horizon for a mid-term view. An ideal system of global welfare is not imminent, so whether this remains a long-term goal, an inspiring hope or highly improbable in the minds of different humanitarians is not critical to where the system needs to focus its efforts in the next five years. The call for global welfare is an important motivating vision,

but the system and international society are still some distance away from such a goal. Small steps towards improvement may be more useful than grand aspirations for the people agencies can actually reach. Focusing expectations in this period means concentrating on what is feasible. This will require four main steps.

First, the system requires a realistic appraisal of its current limits and the external pressures and internal weaknesses which continually compromise it. In particular, every humanitarian programme must be designed and evaluated within its particular global and local context of political possibility, and not judged against ideals of humanitarian perfection alone. As we have seen, the application of humanitarian principles like impartiality and independence are more often thwarted in the outer realm of politics than the inner realm of agency competence. Precise, contextualised evaluation will help to distinguish usefully between political failures and humanitarian failures.

Second, within these limits, the system needs to identify the things it really can change and the softer border points where it can push those limits in the next five years. Many of the things which can be changed will be inner-realm organisational things which have already been identified around knowledge, expertise, capacity, effectiveness, efficiency, accountability, collaboration, linking relief and development, and innovation. But there will also be outer-realm changes in political attitude, wider ownership of humanitarian ideals and greater global impartiality where the system could make significant if uneven progress.

Third, developing a set of reasonable five-year expectations within and at the edges of the system's current limits will mean taking a highly strategic approach to improvement, instead of a very open-ended one. Humanitarian leaders could usefully set ten keynote objectives for external and internal change and improvement across the system, and work to meet them together. These objectives must be simply expressed as either process or outcome standards – understood by everyone and easily monitored and reported. Ideally, some will be easily adapted from the Code of Conduct.

Last, to monitor its progress against these keynote objectives, the system needs to set up the kind of oversight mechanism that has the power to evaluate improvements over the next five years. This is the dashboard and control room for gauging success which still does not exist in the humanitarian system. ALNAP could

build on its experience to take up this monitoring role and assess how the collective efforts of the system are progressing. As well as continuing with its regular retrospectives, ALNAP could then also be more forward looking and measure performance against the system's keynote objectives. To be taken seriously, ALNAP would also need some high-level power behind it. An independent high-level performance panel appointed by the system could actively investigate, consider and report on individual and collective progress and performance.

1.4.1   A culture of improvement

The approach taken to this five-year plan will be critical and must generate a culture of improvement. This must be motivating, widely owned and involve praise as well as sanction. In particular, the culture of humanitarian improvement needs to be positive, unbureaucratic and empowering, and should value a mix of increasing professionalism with continuing discovery and innovation.

1.4.2   Balancing consolidation and innovation

Much of the impetus for improvement in the formal humanitarian system tends towards consolidation at the moment. UN humanitarian reform and the several other improvement initiatives are all focused on bringing organisations and approaches into line so that they work more logically and effectively together. Consolidating good practice throughout the system is deeply important and represents the superlative virtue of the manager. But the system also needs innovation if it is to cope with and adapt to the inevitable shortcomings of its political and practical setting.

Innovation is the virtue of the entrepreneur. As the system moves towards conformity, it must still urgently encourage innovation and entrepreneurialism whereby humanitarian agencies can make new discoveries and risk new approaches. To see that this happens, at least two of the keynote objectives need to prioritise innovation and risk-taking in humanitarian politics and practice. If, in the next five years, the humanitarian system has only consolidated, it will have failed to take advantage of new ideas and will not have adapted to new opportunities.

History often shows that progress cannot wait for reform but requires innovation. Recognising that progress is not likely to be delivered quickly and scientifically in an orderly fashion around organisational reforms, political agreements and UN Security Council resolutions can free agencies to be more genuinely innovative. Less time spent waiting for the new organogram from Geneva, or campaigning for the right resolution to be passed in New York could leave more time for shaping an innovative solution on the ground. Such innovation must surely be the hallmark of a dynamic and entrepreneurial agency.

1.4.3 Encouraging positive thinking

Finally, it might also be good to change the culture of how we evaluate ourselves and our system in the next five years. Peta Sandison's work in Chapter 3 is revealing here. It shows that people in agencies seldom listen to and act upon evaluations. One of many reasons for this may be because most people do not like evaluations, which are often seen to bring bad news. What people want is praise and encouragement, not criticism and finger-wagging. Not unnaturally, therefore, humanitarians tend to shoot the evaluation messenger in the various ways that Sandison describes so well. Maybe they are right to do so?

Evaluations are by nature more negative than positive. Innately, they have more interest in problems than success. They are essentially critical in intent and they do not motivate. Instead, evaluations seem to switch people off and do not spur them on. It would be interesting, perhaps, if ALNAP and its agencies focused particularly on success for a while. For the next five years, evaluations could be encouraged to adopt a method of appreciative enquiry instead of the more intrinsically damning scientific model of measurement against objectives which were often rather hasty or donor-driven in the first place. It might do everyone good if evaluators became more positive investigators, asking 'what's worked well here?' The recipients of aid would, of course, be the key targets of these questions. The idea that recipients are important judges must be central to any new form of enquiry. Identifying positive experience should not mean that negative performances should be overlooked. These should still be exposed, but in the brighter light of what worked well.

A commitment to discovering the good news might well change attitudes and energy within the profession without necessarily becoming an exercise in crazed, deluded

optimism. We would probably learn more and then use this learning more if it came to us as praise and encouragement. This would not solve all the political problems of global equity, political favouritism and hard-nosed obstruction. However, it would mean that when agencies are able to hit the ground and work, they would do so better because they feel better – their minds more filled with models of success than with potential criticisms.

An essential part of this positive thinking is to communicate the millions of good things that humanitarian agencies do. They save lives every day, comfort people every day and employ people every day. Each one of these lives is extraordinarily precious. The system should find better ways to talk about these lives and represent them more powerfully in their thinking, decision-making and evaluations.

1.4.4   Changing international society

Humanitarian action is just one of many emerging but contested areas of potential common action in international society today. Trade, human rights, poverty, arms control, HIV/AIDS and carbon emissions are comparable contests. Many of the same NGOs and UN agencies are also at the forefront of activism on these other issues too. The reality of global competition, the emergence of new powers and the inevitable volatility of politics surely require us to have moderate and hedged expectations of progress on all these issues. This need not make humanitarian agencies into gloomy non-improving fatalists. Instead it allows them draw on a vision of the ideal system to focus realistically upon a shared and constructive mid-term programme in which they have reasonable expectations of themselves and others.

Political optimism may be a useful motivator but calibrated operational realism may make for better programming. We can hope for global welfare but are wiser to focus on more immediate mid-term dials. Particular improvements and innovations within humanitarian agencies must remain the priority of progress in the inner realm of humanitarian practice. In the outer realm of politics, the effort must remain on some key mid-term political objectives such as a proper scale of global need, some proven examples of donor impartiality, and a wider international ownership of humanitarian principles. All these incremental steps need to be monitored by a credible and independent authority with powers to investigate and report on progress against such key objectives.

In its aspirations for global welfare in war and disaster, the Western system might also improve its knowledge of the contribution of other informal humanitarian systems operating today, and make room for them in its calculations of international response. It is probably neither desirable nor politically possible to try to integrate these different systems, but it may be possible to learn from them, to influence them and to anticipate them in the increasing number of environmental disasters that look likely to strike the Earth in its current fragile state.

Last but not least, evaluators need to be realistic about the limits and prospects of the system they are judging in the years ahead. Evaluations that are cautiously progressive and not overly optimistic in demanding perfection of global response may help the system to focus on the possible. More user-friendly evaluations may ensure that learning is taken up, and a concentration on context will ensure that evaluations make real-world judgements about what is good and bad, better and worse, possible and impossible in the world as it actually is over the next five years.

References

Egeland, Jan (2006) 'Solidarity, Not Charity' in *To Guide Global Response, Global Future*, No 2, World Vision International.

Young, Helen et al (2005) *Darfur – Livelihoods Under Siege.* Feinstein International Famine Centre, Tufts University, Medford (June).

Flint, Michael and Hugh Goyder (2006) *Funding the Tsunami Response.* London: Tsunami Evaluation Coalition (July).

OCHA (2005) *Humanitarian Response Review.* UNOCHA (August).

Proportion and distortion in humanitarian assistance

Tony Vaux

2.1 Introduction

This chapter reviews the functioning of the humanitarian system in 2005 through the lens of a set of 43 evaluations deposited with ALNAP (the Active Learning Network for Accountability and Performance in humanitarian action). It will explore the relationship between needs and the allocation of humanitarian aid, seeking to illustrate and illuminate questions of distortion and proportion in the humanitarian system.

The issue of proportion and distortion has been identified as a matter for concern in ALNAP Reviews of humanitarian action since 2001. The 2003 Review observed wide variation in the allocation of humanitarian funding during the 1990s, with people in Ethiopia receiving only $2 per head while those in Bosnia-Herzegovina received $116 per head (ALNAP, 2003, p 14). The same Review went on to note that this was not by accident but because of 'distortion' caused mainly by political factors. The 2004 ALNAP Review found that 'inequities in funding continue', with over 80 per cent of requirements already pledged for the Tsunami response compared with only around 40 per cent pledged for Sudan and the DRC (ALNAP, 2004, p 18). It summarised the reasons for these inequities as 'national interest, geopolitical interest and media focus' and noted, as did the 2003 Review, that the issue had not often been recognised by evaluators.

Until now, the issue of proportionality has not been the main theme of an ALNAP Review. In 2003 the Review focused on Southern Africa and Afghanistan and in 2004 on Darfur, Sudan. The Indian Ocean Tsunami Disaster might have been the obvious focus for 2005, but, instead of looking at the response itself, the current Review examines the implications of the response, and the wider context. The overwhelming scale of funds available for the Tsunami response has finally attracted the direct attention of evaluators. Both the DEC Summary (Vaux et al, 2005) and TEC Synthesis report (Telford and Cosgrave, 2006) give the issue considerable importance, arguing that ways should be found to divert funds from the most heavily funded disasters to those that are neglected.

The underlying principle of humanitarianism is that aid should be allocated according to need. This is the first and most fundamental principle of the Red Cross Code[1] and is widely reflected in the mandates of aid organisations. But, in practice,

aid is subject to political influence and the media. The architecture of aid has taken a particular shape over the years and now reflects its own vested interests. At the level of projects, aid flows through managers and aid workers who have biases of their own and are susceptible to local pressures. At each level there are 'distortions' of the humanitarian principle.

This review seeks to identify key areas of distortion and to examine how they affect humanitarian practice. This analysis is divided into three sections (2.2–2.4 below), on: international political and media influences, structural issues and local interests. The evaluation dataset offers the opportunity to examine what actually happens in the humanitarian system, rather than repeating the general analysis of current trends (of which Vaux (2006) gives a recent summary). This review examines what happens when the shortfall in funding is greater or lesser. For example, high levels of funding tend to create pressures to spend fast, through an operational style, while lower levels of funding lead to more measured and participatory responses. Similarly, the review will explore how aid structures and local interests cause distortions in relation to needs. The food-aid 'lobby' tends to operate against the logic of cash distribution, while a gender focus may lead to increased aid for women and girls when help for boys and men may be equally necessary.

ALNAP's dataset of evaluations for 2005 consists of 43 evaluation reports that agencies have chosen to publish or place on the internet (Table 2.1). The dataset represents only a fraction of the total evaluations conducted and may reflect some biases. Where agencies make a commitment to publish all evaluations, this may inhibit some of the commentary. Others select only those evaluations that they are willing to publish. If this review needs to establish a wider generalisation, reference is made to other studies from a similar period.

In most evaluations, global imbalances are taken as given, even if they arise from serious 'distortions' in aid allocations. Some evaluators have been willing to challenge biases that arise from aid structures, notably in the case of food aid, but the majority of evaluations, such as those by WFP, tend to accept the parameters and simply examine programme implementation. This gives the evaluations little significance for outsiders. Many of the most important lessons arise when the evaluator moves 'out of the box' and takes a broader view of the task. This raises questions about the role of evaluators in examining the wider context, especially in conflict situations where the context may change radically from the time of planning or where the original context was poorly understood.

Table 2.1 ALNAP evaluation dataset 2005: locations and issues

	DONOR	UN	WORLD BANK	NGO[2]	TOTAL
AFRICA	**8**	**1**	**1**	**3**	**13**
	Burundi (2)	Zambia	Mozambique	Niger	
	DRC (2)			Sierra Leone	
	Ethiopia (3)			Zimbabwe	
	Somalia (1)				
ASIA	**5**	**5**	**1**	**7**	**18**
	Afghanistan	Afghanistan	Bangladesh	Afghanistan	
	India	China		Bangladesh	
	Korea	Iraq		Iraq	
	Pakistan	Nepal		Korea	
	Sri Lanka	Pakistan		Pakistan	
				Tsunami (2)	
LATIN AMERICA	**2**			**2**	**4**
	Caribbean			Haiti (2)	
	Haiti				
ISSUE-BASED	**2**	**5**	**1**		**8**
	IDPs	Gender	Natural disasters		
	Finnish assistance	Humanitarian response			
	Food targeting	Pilot food projects			
		Impact on UN reform			
TOTAL	**17**	**11**	**3**	**12**	**43**

Sadly, relatively few organisations actively disseminate their evaluations. ECHO, UNHCR, WFP and the World Bank are notable exceptions. The proportion of NGO (and Red Cross) evaluations that are published is clearly very low indeed, and much lower than among donors and UN agencies. This is a matter of concern, especially given that NGOs tend to favour transparency and accountability as general

principles for their local counterparts. The sources of the 43 ALNAP 2005 dataset evaluations are:

- Donor, 17 (ECHO, 11; Finland, 1; joint, 5)
- UN, 11 (UNHCR/WHO, 3; UNHCR, 1; UNICEF, 2; OCHA, 1; WFP, 4)
- World Bank, 3
- International NGOs, 12 (CARE, 2; Friedrich Ebert Stiftung, 1; Oxfam, 3; SCF, 1; joint, 5)
- National NGOs, 0.

This study has several limitations. The annual crop of evaluations inevitably reflects a degree of chance. The relatively low proportion of NGO evaluations makes it difficult to assess the impact of global distortion at the 'middle' level, while a total absence of evaluations from 'Southern' organisations makes it difficult to nuance the analysis on the ground. On the other hand, several of the current evaluations are themselves syntheses of many other studies, notably Borton et al (2005) and IEG (2006), and in the case of joint evaluations may cover the activity of many different organisations with input from different levels. It is fortunate that this year there have been several major comprehensive studies, including a UN review of the humanitarian system, a synthesis of evaluations relating to internally displaced persons, and major studies of natural disasters by the World Bank.

2.2 International political and media influences

2.2.1 Political influences

One of the key documents in the current dataset is the Humanitarian Response Review conducted by the UN as an overview of the whole humanitarian system. This draws attention to issues of proportion and distortion: 'The uneven support given to forgotten emergencies or neglected needs present the humanitarian organizations and the donors with the dilemma of the equity of their response' (Adinolfi et al, 2005, p 57). It points out that Iraq and Afghanistan continue to dominate global

humanitarian assistance out of all proportion to the needs, while disasters in Africa receive proportionally much less.

This is surely a serious matter. If the objective of the humanitarian system is to address needs in proportion to their severity, then it is clearly failing in its task. The UN report puts the responsibility squarely on the donors who, in turn, have increasingly acknowledged the problem. The Good Humanitarian Donorship initiative, launched in 2003, makes the issue of need-based allocations a key concern. The UK Department for International Development (DFID) has set targets for reducing the gap between responses to different UN appeals.[3] The question is whether such good intentions are likely to be reflected in practice.

ActionAid (Cosgrave, 2004), Christian Aid (Christian Aid, 2004) and Oxfam (Oxfam, 2000; 2003) have all published papers on imbalances in humanitarian aid and the issue of 'forgotten emergencies', but none has given the issue high priority or risked a direct challenge to either institutional donors or the Western public. The issue has received surprisingly little attention from aid workers and is only rarely remarked on by evaluators.

There are institutional and self-interested reasons for this silence, but there are also some reasons that are more obviously valid. The overall level of global allocations for humanitarian assistance has continued to rise, and so countries that receive relatively little may still receive more than in the previous year. This is often the point of reference for aid workers and evaluators, rather than the overall pattern. Humanitarian aid for Africa has increased to record levels (Development Initiatives, 2005, p 8). In 2003, Ethiopia rose to second place in the table of humanitarian assistance, behind Iraq but ahead of Afghanistan, and with a greater amount of funding than ever before (Development Initiatives, 2005, p 9).[4]

Moreover, to some considerable extent humanitarian assistance for Afghanistan and Iraq has been additional (Development Initiatives, 2005, p 27), or more exactly derived from general reserves that had not been allocated to a specific use. Accordingly, it is difficult to quantify the impact on other budgets. It is also difficult to specify the effects of diversion of staff and management capacity. If they had not been busy with Iraq and the Tsunami Disaster, would donors and NGOs have done a better job in Darfur and DRC? This seems likely, but is difficult to quantify.

Data relating to imbalances in global humanitarian assistance are readily available, in the publications of Development Initiatives (2003; 2005). Currently, the Global War on Terror appears to be a major source of imbalance, with excessive focus on Afghanistan and Iraq. Political factors have always influenced donor priorities to a greater or lesser extent. This is not a new problem but it has taken a specific form. The Cold War caused humanitarian aid to be closely linked to political interests, especially in Central America. Today's security interests focus on Islamic areas of the world.

Finally, although it appears that humanitarian aid may be increasing at the expense of development aid (Development Initiatives, 2005, pp 5, 15), it is difficult to specify what has been lost in terms of development. There has been no campaign or protest about this, perhaps because it would be difficult to do so without seeming to question the need for more humanitarian aid.

Because humanitarian action in most locations receives more funding than in the past, aid workers and evaluators either remain unaware of the imbalances or accept them as given. For these various reasons, the problem is 'invisible' and aid workers become conscious of it only in spectacular cases where funding rises to an extraordinary level, as in Kosovo, Iraq and the Tsunami Disaster. Aid workers may also be wary of the issue because it seems to be intractable. Although the Western public may have noted that only certain disasters result in public appeals, it receives little information about imbalances in the overall system. In the case of the Tsunami, agencies that tried to limit or divert public generosity elsewhere were largely unsuccessful in their efforts (Telford and Cosgrave, 2006, p 86). The problem is deeply rooted in Western political systems which link humanitarian aid to other political objectives, and in public responses that are based on images rather than understanding.

2.2.2 The security bias

Political interests translate on the ground into different notions of security. The dominant Western view gives priority to global security over human security, and this results in a tendency to use humanitarian aid as a way of stabilising countries rather than providing assistance according to needs. The problem is illustrated in reports from the ALNAP dataset on Somalia, Sri Lanka and Iraq.

The Somalia Aid Coordination Body was created more than ten years ago and includes most of the key aid actors in Somalia. It works on the basis of consensus and has produced a code of conduct and guiding principles. It provides general information and has developed strategies for response in different sectors. But according to a recent review (Wiles et al, 2004), there remains 'a lack of an over-arching strategy for international assistance to Somalia'. There is no agreed analysis of the various causes and factors relating to conflict. Instead donors have adopted a 'global security' approach and plan their assistance around the notion of a 'peace dividend', meaning that they allocate aid to those places and peoples that are peaceful. Aid is allocated as a reward for achieving stability rather than according to the principle of need. By implication, aid is denied or reduced in those areas that are unstable, despite the likelihood that they may have greater needs or might be areas in which aid could be used more effectively.

The Somalia review notes that there is no information available to assess the effectiveness of this strategy; it is not clear whether it acts as an incentive to other areas to improve, or simply creates division and resentment, which conceivably might increase the level of conflict. The review does find, however, that this strategy is not well received by other stakeholders including NGOs: 'there is widespread misunderstanding about what the peace dividend approach means and considerable scepticism about its appropriateness' (Wiles et al, 2004).

As the evaluation notes, the problem revolves around the lack of any clear idea about the nature of the conflict and the effectiveness of different strategies. The 'securitisation' approach seems to have arisen as a general 'theory of change' rather than from a detailed analysis of the Somali context. A review of 'Aid, Conflict and Peacebuilding in Sri Lanka 2000–2005' (Goodhand and Klem 2005) identifies a similar donor strategy in Sri Lanka, also based on the idea of a 'peace dividend'. Incentives to stop fighting take precedence over immediate needs. The review describes this as a 'conditionality' approach and goes on to observe that 'the lesson about peace conditionalities is that, applied crudely and without a strong political process to back them up, they have limited or even perverse impacts' (Goodhand and Klem, 2005, p 14).

The Sri Lanka review observes also that this approach bears little relationship to analysis of the actual situation. It argues that such strategies have never been proved to work, and are even less appropriate after the Tsunami Disaster where, in an 'over-funded' environment, aid cannot be used effectively as a bargaining tool. The review

accepts the notion that donors should use aid to reduce conflict but argues for a much more nuanced strategy of 'engagement' based on deeper and more sensitive analysis: 'Becoming more conflict sensitive necessarily means becoming more political, in the sense of being more attuned to the political context and governance structures within Sri Lanka' (Goodhand and Klem, 2005). But for busy donors this is a lot more difficult and time-consuming than simply directing aid to those areas that are stable.

It would be a very limited form of humanitarianism that focused only on the outcomes of conflict rather than its causes, but in the current context any attempt to 'manage' conflict is susceptible to political pressure. In Iraq and Afghanistan the global-security agenda is at its most explicit and pervasive, and the donor policies identified in Somalia and Sri Lanka are taken to extremes. Humanitarian action is controlled by joint bodies in which military elements are dominant. Aid is used explicitly as a tool of foreign policy. Because NGOs depend on Western governments for their security as well as their funds, they can make little use of humanitarian principles of impartiality and neutrality.

An evaluation of the NGO Coordination Committee in Iraq (NCCI) (Hansen, 2004) draws attention to the consequences. The behaviour of international NGOs negates 'the appropriate distance between NGOs and authorities or combatants' (Hansen, 2004, p 8). The evaluation bravely warns international NGOs that they are setting inappropriate models of the NGO concept for nascent Iraqi NGOs and the new Iraqi authorities. Instead of contributing to the idea of democratic pluralism they are modelling a form of governance in which NGOs lack independence. This might have a long-term negative impact on governance.

The report urges the NGOs to make strenuous efforts to abide by the Red Cross Code and finds that lack of principle leads to severe problems of coordination.

> In highly politicised contexts such as Iraq, the inevitable differences of principle and philosophy within the humanitarian community become more apparent and emotionally charged, resulting in diminished readiness of agencies to share information, engage in joint planning or cooperate in other ways. In Iraq, relatively more agencies have opted for a 'go-it-alone' approach and have situated themselves outside of coordination structures. (Hansen, 2004, p 10)

The Iraq evaluation shows how politicisation tends to fracture the humanitarian sector and undermine principle. The same process occurs on a global scale. Civil servants in aid departments may sympathise with the notion of impartiality but may not have the freedom to dissociate humanitarian aid from the Global War on Terror. The result is a discrepancy between attempts to strengthen principle, such as the Good Humanitarian Donorship initiative, and clear statements about the prime importance of global security coming from top political leaders.

Among NGOs the line of division falls between those that accept the global-security agenda as their own and those that try to distance themselves from it. Although most NGOs have re-emphasised the principles of impartiality and neutrality, others have taken the view that they should support their country even in its political objectives. This 'Wilsonian' approach to humanitarian aid is most evident in the USA, and has caused considerable tension within internationally federated NGOs, such as Save the Children (Stoddard, 2003).

The general implication is that aid actors need sophisticated and shared analyses, especially when they engage in conflict situations. They need to analyse not only the role of local actors but also that of Western governments which may be funding the humanitarian effort. Aid should be analysed as part of the problem as well as part of the solution. If donors then wish to trump the analysis with global-security concerns then at least this will be explicitly understood. Without being a lot smarter about the interests that surround it, the humanitarian system is, as Fiona Terry observes, condemned to repeat the mistakes of the past (Terry, 2002).

Box 2.1	Evaluation and conflict

Conflict often has severe humanitarian consequences and therefore aid actors want to know how their programmes relate to conflict. But the most common bases for evaluations, such as the DAC criteria, provide little explicit guidance for conflict situations. There is no established method for assessing whether responses are 'relevant' to conflict. In practice, the evaluator has to make an analysis of the context but is rarely given the time or the remit to do so.

CONTINUED

Box 2.1   Evaluation and conflict *continued*

A joint evaluation of donor responses on Afghanistan illustrates this deficiency. It provides only a short history of the country and offers very little analysis of the ongoing conflict. Although the report contains many judgements about relevance, these are not set against an analysis which explains the conflict. It is therefore impossible to say on the basis of this evaluation whether the responses exacerbate conflict, reduce it or 'do no harm' (Danida Evaluation Department, 2005, pp 37–39).

Nearly half the evaluations in the ALNAP dataset relate to countries experiencing conflict, but only 2 reports of the total 43 systematically examine the impact of the humanitarian response on conflict. One of these is not an evaluation but rather a multi-donor study, 'Aid, Conflict and Peace-building in Sri Lanka' (Goodhand and Klem, 2005). This examines donor responses in Sri Lanka on the basis of a strategic conflict assessment conducted in 2000 and now being updated. It provides a detailed analysis of how and where donor activity interacts with conflict. The report shows that conflict analysis can play an important role in evaluation but must be resourced in proportion to the complexity of the issues.

In the second report, CARE evaluates its rights-based approach (RBA) as a way of responding to conflict in Sierra Leone (Napier, 2004). The study begins with a short presentation of the root causes of the war based on the literature. It finds that, in general, the use of RBA has been positive in relation to conflict. But CARE had to select certain 'rights' for focus within the RBA framework, while others were given lower priority. Equity, empowerment and opposition to discrimination and violence were considered to have been particularly relevant. But the analysis was conducted only at the evaluation stage and the programme choices seem to have been largely intuitive. As the CARE study concludes, 'The design of a rights-based project needs to be grounded in a thorough analysis and understanding of the local social, political and cultural context' (Napier, 2004).

The ALNAP dataset shows that evaluators often have to make sense of a conflict response by making a retrospective conflict analysis. It would clearly be better if this had been done at the outset and the evaluator left to check it, but agencies often underestimate the complexities of working in conflict situations. This issue has become more serious for agencies because their funding often comes from governments concerned with 'global' rather than 'human' security.

For evaluators in conflict situations, the problems are even greater than for the agencies. The evaluators are rarely provided with any explicit analysis and yet they have to assess the appropriateness and relevance of the response. In many cases they skim over the analysis, and focus on delivery. But without analysis of the context, judgements about delivery may be of little real value (Box 2.1).

2.2.3   Media influences

The Tsunami attracted attention as both a spectacular type of disaster and because of the deaths of thousands of Western tourists. Fundraising was further boosted because the disaster occurred at a time when many families were celebrating Christmas (Telford and Cosgrave, 2006, p 83). The level of funds raised by the Disasters Emergency Committee (DEC) in the UK was more than five times as much as the highest previous appeal (Kosovo), which itself was considerably higher than the next most successful appeal, for the floods in Mozambique, where a significant reason for the success of the appeal was said to be the image of a woman giving birth in a tree. In terms of public response, humanitarianism is a lottery. The problem for humanitarian charities (or NGOs as they prefer to be called) is that they depend on public sympathy and are therefore reluctant to challenge the pattern of public concern. This gives the media a dominant role in shaping public sympathy and responses.

Political and media interests may combine to generate a public response. The Kosovo crisis was relatively small as a humanitarian event but because of political involvement the media observed the unfolding of events, witnessed the human suffering and had ample opportunity to project it to the public over a long period. In effect, political and security interests become converted into public interest, and this was further converted by the aid agencies into humanitarian responses. In recent years the Global War on Terror has given this a particular twist, with disproportionate levels of funding for areas of highest media coverage, such as Afghanistan and Iraq, and other areas with little political interest being neglected.

The pattern of donor-driven responses can be extended through the system even to the local level. Within the Democratic Republic of the Congo (DRC), the Goma and Ituri crises attracted funding out of proportion to the general level of needs because they happened to receive international media coverage (Borton et al, 2005). Local

media can also distort the response. For example, Oxfam reports in relation to floods in Bangladesh that 'Unfortunately, most of the humanitarian emergency responses were triggered by the media coverage, both for Oxfam and other agencies. Some of the districts did not get media attention, which delayed humanitarian community and donors' response' (Oxfam Evaluation Team, 2005, p 9).

The Oxfam Evaluation Team also finds that 'The situation could be improved by developing the staff analysis and reporting skills'. But, as Darcy and Hofmann explain in their 2003 study for the Humanitarian Policy Group, the notion of needs is not as simple as it first appears, and the art of needs assessment is remarkably slippery. There is little agreement among humanitarian actors on the nature of 'needs'. Very often 'needs' are narrowed down to a choice between the range of items and services that agencies can most easily supply. Methodologies for needs assessment vary so widely that comparisons within and between countries are impossible. Darcy and Hofmann conclude that, in practice, 'Most management decisions about humanitarian response are made primarily on the basis of non-formal assessment' (Darcy and Hofmann, 2003, p 25).

In the Bangladesh case, media pressure was translated to Oxfam through the offices of donors, and spelt out in considerable detail:

> The Oxfam flood response was donor-based and it was designed with donor money in mind, and this can detract from a needs based approach. This approach has influenced the number of families to be covered in a district regardless of the severity of the disaster. (Oxfam Evaluation Team, 2005)[5]

The Oxfam evaluation argues that this compromise was unnecessary: 'Oxfam being an organisation which can attract large amounts of funds immediately, internally and externally, should concentrate on covering more affected areas rather than adopting a more fund/donor based approach' (Oxfam Evaluation Team, 2005). And the report goes on to urge management to take a more robust position in relation to such pressure.

This unusually frank analysis of practical realities illustrates how easily humanitarian responses can be swayed by what amounts to little more than rumour and personal perception. Instead of commissioning assessments based on common approaches, aid managers compete with each other to offer help in high-profile areas

where the media and politicians are likely to congregate. In the absence of a shared system for assessment, humanitarian outcomes depend on processes of negotiation in which a host of issues arising from agency interests and structures take precedence. Or, in other words, the 'need' is largely a construct of different interests and priorities among all the different stakeholders.

Consequently, an agency that takes considerable trouble to establish the facts does not necessarily gain funding support. The problem emanates largely from the donors, but UN agencies and NGOs seem to be defenceless against it. The increase in donor funding for humanitarian purposes in relation to development has led many international NGOs to allocate their un-earmarked funds for core costs and development, and pare down their humanitarian reserves to a minimum.[6] As NGOs receive a greater share of donor allocations, the budgets of UN agencies have been squeezed (Macrae et al, 2002, pp 3–4). Donor funding remains almost entirely earmarked for specific purposes rather than allocated to general funds that can be drawn upon according to need.

In the Humanitarian Response Review this is noted as 'a major constraint' (Adinolfi et al, 2005, p 55). Both the UN and NGOs are very susceptible to donor pressures, and the tendency has been to compete rather than cooperate. In the absence of any concerted effort on the part of these 'implementing agencies', allocations of aid will continue to follow the lowest common denominator of political interests, colonial ties, media reports and the personal views of managers. It may be difficult to define what 'need' really is, but it is usually fairly easy to recognise when other factors are predominant.

2.2.4   A case of high-level funding: the Tsunami

The DEC evaluation of the Indian Ocean Tsunami response observes that:

> The Tsunami Disaster has thrown up a challenge to the whole humanitarian system. A massive response to one disaster is excellent in itself but agencies are not geared up to respond without taking staff resources from other disasters. How can the DEC ensure that the success of one public appeal does not have negative effects in other parts of the world? (DEC, 2005, p 7)

Although the Tsunami Disaster generated massive additional funding, the capacity of the humanitarian system is not entirely elastic. In order to mount such a rapid large-scale response, staff had to be drafted in for the Tsunami response from other needy areas. Moreover, senior management and supporting departments had to devote much of their attention to the Tsunami, neglecting other areas. This lack of 'surge capacity' in the humanitarian system has also been noted in the TEC Synthesis evaluation (Telford and Cosgrave, 2006, pp 117–118) and features as a main finding in the UN Humanitarian Response Review (Adinolfi et al, 2005, pp 39–42). The conclusion is that if any single disaster receives disproportionate attention, it is likely to undermine the response elsewhere.

Both the DEC and TEC evaluations observe that the high level of funding for the Tsunami Disaster, together with its high profile, had effects on the form and content of the response. Some of these effects were positive, giving agencies the resources to scale up and experiment. But there were also negative effects. First, according to the DEC and TEC evaluations, the relief phase was extended unduly because agencies were to some extent 'commodity-led'. So long as they had materials coming through the pipeline to distribute, they continued to distribute them, regardless of changing needs on the ground. Second, pressure for speedy results combined with ample availability of funds led to a more operational style of response with greater reliance on expatriate staff, and a tendency to overlook local capacity. This problem, noted particularly in the TEC Capacities evaluation (Scheper et al, 2006), led to a third problem – a difficulty in moving from relief into recovery mode because relationships with local organisations and local personnel had not been developed during the relief phase.

The TEC Synthesis report concludes that the Tsunami response represented 'a missed opportunity' characterised by 'Supply-driven, unsolicited and inappropriate aid, and inappropriate housing designs and livelihood solutions' (Telford and Cosgrave, 2006, p 93), and that 'Allocation and programming, particularly in the first weeks and months of 2005, were driven by politics and funds, not by assessment and need' (p 20). This lack of proper assessment was also a reason why such a high proportion of funding was used for relief, and the transition to recovery was slow (Telford and Cosgrave, 2006).

Two main conclusions emerge from this. First, an extraordinary response to one disaster will have negative impacts on other disaster responses, even if additional

funds are available. It is impossible to specify these exactly but the crucial elements are transfer of experienced staff and diversion of senior management attention. Second, if there is undue pressure to achieve profile or spend funds rapidly, the quality of the humanitarian response will suffer. Accountability to donors takes precedence over accountability to beneficiaries. In the case of the Tsunami, the international response was effective in terms of relief, but much less so in terms of building local capacities and reducing future vulnerability. Having failed to address these issues in the first year, it will be difficult to tackle them later. It is not that aid agencies always ignore the need to develop local capacity, but that a very high level of funding and high pressure for quick results establish a different order of priorities.

2.2.5   A case of low-level funding: the DRC

The Tsunami Disaster caused over 200,000 deaths, but insecurity in the Democratic Republic of Congo (DRC) is thought to have caused 3.9 million deaths between 1999 and 2004 and more than 2 million people have been displaced from their homes on a long-term basis.[7] According to FAO, the DRC has the highest level of food insecurity in the world, with 75 per cent of the population undernourished (Michael et al, 2004, p 40). Many of the deaths occurred through the indirect effects of war, such as poor health and nutrition. By almost any measure, other than political interest and media coverage, DRC should have been the main focus for humanitarian attention in the last decade.

But this was far from the case. The TEC Synthesis report estimates that international pledges and donations for the Tsunami Disaster exceeded the estimated total economic impact of the Tsunami ($9.3bn) and provided more than $7,100 for each person affected (Telford and Cosgrave, 2006, p 20). There are no reliable estimates for numbers 'affected' in the DRC and therefore it is difficult to calculate the level of aid per head but probably this was in the region of $5–10 during the height of the war.[8] Total humanitarian assistance provided to DRC in 1995–2001 was less than 10 per cent of what was provided for the Balkans in the same period.[9] Such figures provide only inexact comparisons, but, whatever method is used, the conclusion must be that people who were affected by the world's worst tragedy of its time received very much less help than those who were affected by disasters that attracted international media attention.

Are there any positive aspects of this low level of funding? Not many. Lack of funding led to unnecessary deaths. But it may be worth asking whether the response was different from that to the Tsunami Disaster. Two evaluations of ECHO's response in DRC confirm that lack of funding had serious negative effects, but also indicate that in small ways agencies were able to make a virtue of necessity. A general study of ECHO's interventions notes that a lack of implementing agencies of sufficient strength was the main constraint on ECHO extending its programmes (Michael et al, 2004).[10] European NGOs are very few on the ground and, according to this report, 'most partners have reached (or overstepped) the limit of their operational capacity' (Michael et al, 2004, p 25). As a result, both ECHO and its partners in DRC have to operate a kind of triage, choosing between desperate unmet needs and sometimes having to make decisions according to capacity rather than need. ECHO has gone about this in a systematic way, dividing the country into zones based on vulnerability (ECHO (2006) gives the methodology). This makes it possible to prioritise certain areas, although the distinctions are often marginal and the information uncertain. Those areas that are excluded would probably rate very highly in almost any other context.

The response in DRC is hampered by shortage both of resources and capacity. The ECHO evaluation argues that, within DRC, priority should be given to quality of response in specific locations rather than geographical extension. But in the absence of other funding this may mean that other regions receive no aid at all. For example, 'partners face the dilemma between the developmental strategy of supporting a complete health system, and allocation of limited means' (Michael et al, 2004). In practice, aid managers in DRC have to make life-and-death decisions – an especially unfair burden when managers in other situations may be worrying about how to spend funds fast enough.

Even within the prioritised zones in DRC, the level of funding is inadequate for basic humanitarian tasks. Managers have to decide which group, among many in desperate need, should be prioritised above another. There can be no ideal solutions because the funding is simply inadequate.

> ECHO is currently spending US$2.64 per capita/year in the zones its partners are covering. A World Bank estimate that pitches the cost of delivery of basic health services is US$12 and a more recent WHO opinion is that below US$10 'no equitable nor effective health care is realistically within the reach of the health sector'. (Michael et al, 2004)

To make matters worse, the costs of delivery are relatively high in DRC because of the effects of war, and the condition of the population is extremely poor, increasing the necessary scale and costs of health and nutrition services.

> The classical 'emergency reasoning' that children under five and women in fertile age are the most vulnerable and therefore the only target group for nutritional intervention does not apply in the particular situation of the DRC. The protracted emergency has compromised population groups – adolescents and adults – that are not traditionally considered high risk and high priority. (Michael et al, 2004, p 47)

Lack of funding means that user fees have been introduced and then increased in order to support health services. These are charged at a 'realistic' rate, with the inevitable risk that very poor people may be excluded. Funding dictates who gets medical care and who does not. When fees were raised between April and October 2004, rates of use fell by 60 per cent (Michael, 2004). User fees may have some advantages in situations where they serve to encourage local engagement and ownership of local health services. But in DRC they simply exclude large numbers of very needy people.

While there was little that agencies could do to save lives after the Tsunami Disaster, in DRC aid could have saved lives. More money for health services would have enabled poor people to gain access. Objectively, aid agencies should be massing in DRC even today, although the worst of the disaster has passed, and should have been rather lightly represented in the aftermath of the Tsunami. But the reverse is the case – and very markedly so.

In practice, the disastrous situation in DRC was treated as a problem of under-resourced development rather than as a humanitarian crisis in which life-saving and minimum standards created an absolute moral responsibility. This may reflect a more general tendency to switch to development mode ('do what you can' rather than 'do what you must') as a way of avoiding humanitarian responsibilities in those situations that fall victim to the pattern of imbalance and distortion.

This represents a terrible failure of the aid system, but there are some small positive aspects of the response that deserve attention. On the ground, lack of data (because of poverty and the disruption caused by war) and lack of expatriate staff willing to

work in such insecure areas forced aid agencies and donors to give more freedom for local choices and more responsibility to local staff and organisations. The ECHO evaluation team notes that the programme considered had to rely heavily on information provided by local communities. This forced the programme to focus on building local capacity rather than on direct delivery of services:

> the poorest or most vulnerable families in a given area can only be identified with well-structured methods of community participation. If the project design is to provide a basis for continued and increasingly developmental approaches, a strong emphasis on beneficiary training, partner capacity building, community participation and involvement of the state structure where this still exists, is necessary. (Michael et al, 2004)

If more funding had been available (perhaps because of some spectacularly successful media exposure) DRC might have experienced a massive expatriate-run operational response, as in the aftermath of the Rwanda genocide, or more recently in the case of the Tsunami Disaster. But DRC suffered from lack of both funding and security. Expatriates were reluctant to go there and this may have reduced the media coverage still further. With little money and few expatriates, decisions and responsibilities in DRC were handed over to local communities, whether or not they could cope with them. As a second ECHO report, which looks specifically at the work of its implementing partner GOAL, notes, 'Growing its own timber is the only way forward for a relatively poor and isolated area' (Michael, 2004).

Ironically, lack of funding and profile may have made agencies more compliant with Red Cross Code principles, such as building local capacities and (thereby) reducing future vulnerabilities. It may have led to a closer relationship with 'beneficiaries' and greater willingness to involve them in decision-making as required by the Code. But none of this should obscure the fact that the most basic of all the principles was violated. The first principle of the Red Cross Code, known as the 'humanitarian imperative', demands that aid should be allocated according to need. On a global scale this was clearly not the case.

The contrast of the DRC response with the Tsunami response demonstrates that lack of funding encourages a closer relationship with beneficiaries, whereas too much funding tends to create a tendency to overlook local capacity. In cases where funding is modest in relation to needs, donor agencies are obliged to play a

supporting role in relation to local capacity. Where funding is lavish in relation to needs, donors and implementing agencies assume a dominant role. As the TEC Capacities evaluation (Scheper et al, 2006) observes, this means that local capacities dwindle in a context of heavy-handed funding, and local actors may even become hostile to foreign agencies and refuse to cooperate in longer-term recovery and reconstruction.

Arguably, 'over-funding' tends to work against good practice but 'under-funding' will result in failure to meet fundamental purposes, and tends to push back to the community responsibilities that the community may be unable to sustain. In terms of timescales, 'over-funding' is more effective in its immediate impact but may be negative in the longer term, whereas under-funded responses may have little immediate impact but might be more effective in the longer term, building capacity and reducing vulnerability. Neither of the extremes is desirable.

Much of the good practice in DRC was not deliberate, but forced on the agencies by circumstances. Participation occurred in order to fill a funding vacuum. Pressures that might have led to posturing for profile, competition and expatriate-led responses were absent. Good practice arose from the absence of negative factors. But there is little sense from the reports that aid workers had a clear idea of upholding principle and holding out for good practice in the face of a challenge. The overall picture is of a humanitarian sector that welcomes the charity of donors rather than asserts 'rights'.

2.3 Structural issues

2.3.1 Imbalances between sectors

The most extreme forms of imbalance occur in the geographical spread of humanitarian assistance, and from a humanitarian perspective these might truly be called 'distortions'. But there are other inequalities in the humanitarian system that are less deliberate but arise from the way in which humanitarian structures have developed over the years. Organisations specialise in certain sectors or types of

response, but as the needs or the analysis changes they continue to follow the same approaches, becoming 'vested interests' within the system. This applies not only to specific institutions but also to the influence of historical momentum within all agencies. There is a tendency to do what has been needed in the past, rather than what needs doing now. The problem is not necessarily self-correcting. Strong agencies in particular sectors have the power to attract greater resources; by implication, other sectors may be neglected.

From the evidence in the ALNAP dataset, how great is this imbalance? The UN Humanitarian Response Review seeks to map the capacity of the entire humanitarian system. It finds particular deficiencies in relation to camp management and protection.

It attributes deficiencies in the water and sanitation sector to discrepancies between stated capacity and real capacity (Adinolfi et al, 2005, p 10). UNICEF often takes responsibility for water and sanitation within the UN but has so many other responsibilities that it has not been able to meet expectations. Although NGOs such as Oxfam have developed significant capacity in the sector, there is still a shortfall, especially in sanitation. A plausible reason for the under-representation of this sector is that it is much less photogenic than, say, child feeding.

Shelter has emerged in the responses to the Tsunami and Pakistan earthquake as another neglected sector, with limited institutional capacity within the UN and no international NGO that is considered pre-eminent in this field. The DEC Tsunami evaluation notes:

> A striking feature of the tsunami response has been lack of expertise on post-disaster housing among DEC members. In the absence of a collective approach, this has given rise to an array of house designs which has caused confusion. There is a need for DEC members individually and collectively to enhance their capacity in relation to housing. (DEC, 2005)

Lack of technical expertise in the shelter sector has left agencies open to the pressure for rapid results. Shortly before the anniversary of the Tsunami Disaster there was a rush to start housing projects in order to demonstrate 'success' to the media. Many of these were 'cherry-picked' as the easiest but not necessarily most important projects, and were largely implemented by contractors rather than

through self-help. But as the DEC and TEC evaluations point out, experience over many years clearly indicates that creating permanent housing is a very slow process and it is usually at least a year before the policy environment has been sorted out. Agencies should have worked together to establish a realistic timetable and common plan, but competition won the day. The need for transitional and temporary shelter was largely ignored, and so tens of thousands of people were still living in tents a year after the disaster.

Although the ALNAP dataset includes little material on the Pakistan earthquake,[14] there are some indications that agencies have engaged more closely and imaginatively with local communities in this response and made much better use of self-help approaches. In a recent edition of *Humanitarian Exchange*, Causton and Saunders (2006) describe how Catholic Relief Services (CRS) has used cash grants with local technical inputs for shelter reconstruction, aiming to provide at least a first room which the occupiers could later improve. Although no evaluation is yet available, the project is said to have been successful in engaging with local beneficiaries and has overcome the problem so poorly addressed in the Tsunami Disaster, the needfor transitional housing:

> Households themselves chose to develop what was initially envisaged as an 'emergency' shelter programme into the beginnings of permanent housing, and also took responsibility for site planning, the sale or transfer of assets and the use of salvageable resources…. Trusting in the resourcefulness and ingenuity of affected households themselves can prompt creative solutions to perceived challenges, and result in far greater long-term impact. (Causton and Saunders, 2006, p 12)

If agencies are to improve their ability to respond to shelter needs they will have to consolidate and build on the experience of the Tsunami and Pakistan disasters. Part of the problem may be that the need for emergency shelter does not arise in every disaster. There was a long period in the 1990s when this type of expertise was scarcely needed. Agencies may be reluctant to invest in a capacity that is not always in use. Without a strong institutional base, shelter may continue to be neglected in the humanitarian response.

2.3.2   A dominant sector: food aid

At the other extreme, certain sectors are over-represented in the humanitarian system and tend to cause distortion favouring their own area of competence. One of the most striking examples is food aid, which has retained a pre-eminent position in disaster responses despite ever-mounting evidence that cash responses are preferable. The distorting influence of food aid arises primarily because certain donors, notably the USA and Canada, subsidise their domestic food production and are left with surpluses to dispose of. Although the EU has stopped using food aid to dispose of surpluses, there is still a strong tendency for the humanitarian system to favour food aid above other kinds of response.

The policy of food-aid donors is now reinforced by the structures of the humanitarian system. The World Food Programme (WFP) is the largest and most influential UN agency in many disaster-affected countries. This imbalance is not of the UN's making but comes from the donors, and is further exaggerated by their responses to appeals. Smillie and Minear find that 'Although food represents only half of what was requested in most CAPs, it usually received about 75% of donor resources' (Smillie and Minnear, 2004, p 191). So great is the preponderance of food aid in humanitarian responses that a number of NGOs in the USA are reported to have become dependent on food aid as a source of income and have also developed vested interests in the continuation and expansion of food aid (Barrett and Maxwell, 2005).

The problem is further exacerbated by the way in which humanitarian needs are assessed, through the Consolidated Appeals Process (CAP) of the UN. The deficiencies of the CAP (and its variant the Consolidated Humanitarian Appeals Process or CHAP) have been examined in a number of studies, and are now widely recognised (Darcy and Hoffmann, 2003; Smillie and Minear, 2004).[12] In short, the CAP is little more than a list of the requirements of aid agencies and is almost exclusively focused on the UN. Although donors tend to refer to CAPs as the main measure of need, they are actually based mainly on bids from UN agencies according to what they think they can do rather than the totality of need. As a substantial proportion of the response, including practically all the work of NGOs, takes place outside the CAP, it cannot logically be viewed as a complete estimate of need but is often used as such in the absence of anything else (Smillie and Minear, 2004, p 191). This means that donors base their aid allocations on supply rather than demand.

As stated in the CAPs, 'needs' are likely to represent the relative size and capacity of UN agencies rather than anything else. Food aid often becomes the top priority because WFP is the largest and most active agency. By contrast, agricultural inputs are typically under-represented and, with less capacity in FAO, likely to be under-funded because donors lack confidence. Even those donors which do not need to dispose of food surpluses sometimes subscribe to the 'food aid mindset' of UN appeals. An evaluation of responses to the Ethiopia emergency of 2002–3 concludes that 'despite the escalating crisis, the donor response to the non-food requirements in the appeal remained largely stagnant until March 2003. There was an initial lack of appreciation that the crisis was far more complex than food availability' (Steering Committee, 2004, p 25).

The humanitarian system has not moved far from the 'food availability decline' model of famine which Amartya Sen debunked in 1981 (Sen and Dreze, 1981). It is still common to see calculations for food shortages translated directly into food-aid requirements, without consideration for local purchase or cash inputs that might stimulate local markets. Sen showed that famine was a result of falling entitlements rather than lack of food in the markets, and the implication was that distribution of cash or other entitlements would be the logical response.

It has taken the humanitarian system a quarter of a century to come to grips with this proposition. The notion of cash inputs instead of food aid is now receiving considerable attention.[13] The ALNAP dataset includes several papers that support the need for cash inputs rather than food aid. One evaluation in Ethiopia quotes research conducted by the Ethiopian Economic Association and the Ethiopian Economic Policy Research Institute of a Save the Children 'Cash for Relief' (CfR) programme to show that food bought with cash aid is 39 per cent cheaper than imported food and 34 per cent cheaper than grain purchased from national markets in Ethiopia. Moreover, the price at which an aid agency might buy grain on the local market is 7 per cent higher than the price paid by CfR beneficiaries (Brandstetter, 2004, pp 19–20). The same report finds that the pilot programmes were an 'unqualified success' (Brandstetter, 2004, p 22), and asserts that 'The evaluator knows of no other relief intervention that is more effective in both preserving and rebuilding assets than CfR'.

Another USAID study in Ethiopia (Gregg, 2004) focuses on distributing cash for seed, comparing this with distribution of seed, as in previous programmes.

It similarly concludes that 'cash infusions stimulate the local economy and bring poor farmers into it. Purchasing seed empowers the farmer and stimulates decision-making'. If cash inputs were used more widely it is possible that opposition to food aid might develop and the need to choose between accountability to donors and accountability to beneficiaries might fall directly on the implementing NGOs. An evaluation of cash-based programmes for Oxfam in Haiti found that 'The project beneficiaries were particularly opposed to the distribution of rice as part of the cash for work remuneration; they wanted to make their own financial decisions and decide what food variety to purchase' (Creti, 2005, p 4).

But is USAID concerned about 'accountability to beneficiaries'? This is by no means clear. The findings of the evaluations in Ethiopia are consistent with a wider review of 'Cash and vouchers in emergencies' by Paul Harvey for the Humanitarian Policy Group (HPG). He concludes that 'the existing documentation of cash and voucher-based responses shows that they are overwhelmingly successful in terms of their impact' (Harvey, 2005, p 3), but

> Given the arguments in favour of cash-based responses, why have agencies remained so reluctant to use them? The way in which the architecture of the humanitarian system is currently structured seems to inhibit consideration of cash and voucher responses. In the UN system, in particular the consolidated appeals, the almost complete absence of cash or voucher-based approaches suggests that cash is not seen as an option in part because the dominant operational agency is mandated to provide food. (Harvey, 2005, pp 3–4)

The food-aid lobby within the UN is powerful enough to 'distort' the entire humanitarian system. It has taken many years of research even to raise the question of food aid, and so far there has been little change in practice by donors. The UN is susceptible to financial pressures for accountability to donors and only weakly reflects the notion of accountability to beneficiaries. As in the case of food aid, donor interests easily become institutionalised. The UN General Assembly does little to represent the beneficiaries and so the tendency is for the aid architecture to represent powerful donor interests and specific interest groups rather than global 'need'. Change comes about painfully slowly, if at all.

2.3.3   A neglected issue: internally displaced persons

A similar imbalance may be noted between the attention given to refugees and the attention given to internally displaced persons (IDPs). Refugee rights have been clearly demarcated since the Refugee Convention of 1951 but IDPs were only recently brought under international protection through the 1998 UN Guiding Principles on Internal Displacement. A large and powerful UN agency is concerned with refugees, the High Commission for Refugees (UNHCR), while IDPs are represented by an Internal Displacement Division (formerly Unit) managed by the Office for the Coordination of Humanitarian Assistance (OCHA). Donors take a particularly close interest in refugees as potential asylum seekers, but, in terms of humanitarian need, IDPs are probably more important. Typically there are more than twice as many IDPs in the world as there are refugees (IFRC, 2005, pp 218–223). This imbalance, perpetuated in UN structures, causes perverse effects. The strong legal protection offered to refugees and the promise of humanitarian support from UNHCR sometimes draws IDPs across international borders to become refugees.

One of the most important reports in the ALNAP dataset considers the lessons to be learned from evaluations relating to IDPs (Borton et al, 2005). This is a synthesis of 11 main reports, supported by other studies and a process of interviews and discussions. It was commissioned as a joint evaluation by a group of donors and UN agencies. The synthesis report finds that although IDPs lack the clear protective force of international law, the Guiding Principles have now been widely adopted and have helped to improve the status of IDPs. There is an ongoing debate about the responsibility for IDPs within the UN system and there are still some unresolved issues, such as responsibility for camp management within the UN.

In recent years the tendency with regard to IDPs has been to avoid giving responsibility to a 'lead agency' such as UNHCR, but to rely instead on a 'collaborative approach' among UN agencies. This has not always been satisfactory. The UN Humanitarian Response Review finds that the attempt to work by 'goodwill and consensus' depends in practice on the skills and authority of the Resident Coordinator or Humanitarian Coordinator. It concludes that the quality of these coordinators is too varied and urges the development of a cadre with a higher level of authority and expertise (Adinolfi et al, 2005, p 11).

The IDP synthesis report cites a number of other reasons why the humanitarian system is not good at dealing with IDPs. Problems include 'Lack of access, inadequate funding, difficulties in the identification of IDPs and their needs, and assistance not being sufficiently needs driven' (Borton et al, 2005, p 11). In particular, this report describes the lack of needs assessment as 'distinctly unimpressive' and argues for 'comprehensive, multi-sectoral, inter-agency assessments' (Borton et al, 2005, p 16).

This is important because IDPs, along with refugees, are not necessarily needy. A major debate in relation to IDPs concerns whether they should be treated as a separate category at all. This arose because they were often poorly treated in relation to refugees, but the debate has created a new problem – the status of IDPs in relation to the general population. In the words of the synthesis report, 'The evaluations revealed a strong vein of objections, not only to the treatment of IDPs as a separate category but even to their separate identification amongst all actual and potential vulnerable groups' (Borton et al, 2005, p14).

The introduction of the Guiding Principles runs the risk of creating a special category of humanitarian aid that, as in the case of refugees, may ultimately lead to distortion in relation to needs. Unfortunately, the division of the UN into separate agencies tends to encourage the separation of populations into different groups (refugees, children, farmers, etc). This process could go on indefinitely, with new agencies being added to deal with each group. ECHO has already expressed fundamental opposition to treating IDPs as a separate category. This view is supported in several of the evaluations: 'separate identification was at odds with needs-driven humanitarianism…. IDPs are better off than some other vulnerable groups, including those who did not leave their homes in the face of insecurity and threats to their protection' (Borton et al, 2005, p14).

The synthesis report argues that the real focus of neglect is not IDPs but protection: 'there is evidence of a continuing and substantial deficit in the protection work done by the international community' (Borton et al, 2005, p 12). Even for refugees, the level of protection is too low, but the focus on refugees leaves other groups almost entirely neglected. The ALNAP *Review of Humanitarian Action 2004* described the Darfur emergency as a 'crisis of protection', expressing this at two levels: the failure of the international community to exert effective pressure on the government, and the failure of operational agencies to protect the people of Darfur from violence

Box 2.2   The protection deficit

Where national governments fail to protect IDPs, there is evidence of a continuing and substantial deficit in the protection work done by the international community. Practices and omissions contributing to this deficit include:

- Overlooking the protection needs of minorities
- The prioritisation of material assistance over protection needs
- The inability or unwillingness of implementing partners to engage in protection work
- European domestic asylum policy compromising in-country protection work
- Lack of access in areas of insecurity
- Lack of adequate monitoring of human rights abuses
- Conceptual confusion in donor organisations about their potential role in relation to protection
- Inadequate levels of funding being provided for protection activities in some cases.

Source Borton et al (2005) p 12.

(ALNAP, 2004). In response to increasing interest in this issue, ALNAP has published a guide on Protection for humanitarian agencies (Slim and Bonwick, 2005). But although protection is now much better recognised as an issue by humanitarian agencies, the international community has shown little vigour in implementation, especially in the case of IDPs.

The IDP synthesis report concludes that displacement should be 'used as an indicator of potential vulnerability rather than as a means of defining target groups' (Borton et al, 2005, p 15). The difference lies in vulnerability, rather than need. Or, arguably, need would be more usefully defined in terms of vulnerability.

The first principle of the Red Cross Code is that 'need comes first', but it is not clear to what extent this includes vulnerability. Long-term vulnerability is not addressed until the eighth principle where it is seen more as a consequence of disaster ('future vulnerability') rather than a preceding condition (risk) or potentially the

overwhelming issue during an ongoing disaster (protection), as in the case of Darfur. This may reflect an underlying assumption that states are responsible for security, while Western aid seeks to address physical needs. But such a model is not applicable in the case of fragile states, or where the state turns against its own people, as in Darfur.

It is becoming increasingly important to distinguish between vulnerability and needs. The reason why refugees were singled out for special attention was their vulnerability, but this has been transposed into a focus on their 'needs'. A joint UNHCR/WHO evaluation of health programmes for Afghan refugees living in Pakistan finds that the level of services provided and general health status of the refugees is better than that of the host population. But instead of examining whether this is justified in relation to threats to their vulnerability, the report concludes that 'This does not mean that UNHCR is aiming too high, only that the standards of the host country and that of Afghanistan are very low' (Michael et al, 2003, p 25).

This is clearly an unrealistic argument. Donors are not going to raise standards across the whole of Pakistan and Afghanistan (and the rest of the world) in order to match the standards set for this group of refugees. The issue becomes even more complicated in relation to sub-groups of refugees. Among Afghan refugees in Pakistan there is a substantial discrepancy between those living in camps, who receive considerable support, and those living as best they can in urban areas with little support from donors: 'For the latter, housing and living standards are often poor, if not appalling, and primary school attendance is rare – children commonly work as garbage collectors' (Michael et al, 2003, p 25). A focus on protection would seek to ensure that such refugees could live normal lives free of violence, but a focus on 'needs' creates an incentive to live in camps.

The humanitarian system is geared to deal with categories of people. It focuses on people in camps and tends to neglect those who live in scattered host communities. It tends to favour refugees in relation to IDPs, and focuses on needs rather than vulnerability. Over the years, these distortions have become instituted in structures and ways of thinking. By taking the 'camp' situation as the norm, even the Sphere Technical Standards have tended to reinforce this tendency. Aid workers feel uncomfortable in situations where the Standards cannot be applied, even though this may include a very large part of the world. Institutional ways of thinking tend to avoid the difficult situations and issues.

2.3.4   Structural divisions: relief, recovery and reconstruction

As well as a tendency to focus on need rather than vulnerability, the ALNAP dataset also shows that the humanitarian system tends to allocate disproportionately large sums to relief and reconstruction rather than to recovery. In particular the system neglects the importance of livelihoods. This comes out particularly strongly in the TEC evaluations but is also supported by other findings. It reflects a more general tendency to focus on what can be supplied (relief items, houses) rather than what has to be supported through engagement.

It is perhaps interesting to note that 'recovery' is the only term in the sequence of humanitarian interventions (which may also include 'rehabilitation') that implies agency on the part of the affected people. 'They' have to recover, and define what that means, whereas 'we' provide relief, rehabilitation and reconstruction. No wonder, perhaps, that recovery is the Cinderella among phases of the response.

A long-term study of disasters in Mozambique (Wiles et al, 2005) for the World Bank and ProVention Consortium notes that the effects of this bias can be particularly acute where the overall level of funding is low. Donors have a tendency to focus on livelihood recovery only when sufficient funds happen to be carried through from the relief stage and become available for use in recovery, as in Mozambique after the spectacular and highly publicised floods of 2000.

> The 2000 floods in Mozambique demonstrated clearly that it is possible to make an impact and carry out extensive recovery activities when the disaster is high profile and the amount of money donated to the affected populations is large. In lower profile emergencies, all the donated money is usually absorbed during the expensive relief period, leaving little room for maneuver in the recovery period. This was not the case in Mozambique where resources were pledged and continued to arrive over the two-year post-flood period. (Wiles et al, 2005, p 24)

A similar study of disasters in Bangladesh, also for the World Bank and ProVention Consortium, finds that, once the relief phase was over, donors immediately focused on the reconstruction of infrastructure – thereby omitting the recovery stage. Opportunities to link these projects with livelihood support were missed, leading to the conclusion that 'the focus of major donors on infrastructure may be one of the

reasons why the understanding of, and attempts to build on, livelihoods after the floods was patchy' (Beck, 2005, p vi).

Lack of attention to recovery probably reflects lack of representation by 'beneficiaries'. The Bangladesh study indicates that, where they have the chance to express an opinion, beneficiaries emphasise the importance of livelihood recovery rather than relief or reconstruction. NGOs in Bangladesh are well aware of this and have become so well organised and articulate that they have been able to influence government to move from 'flood control', with the emphasis on physical barriers, to 'flood proofing', meaning rapid support for the livelihood strategies of the affected population.

NGOs are able to make use of their extensive savings and credit networks to provide short-term loans. But it is essential that these payments should be made immediately after the disaster, before markets and chains of supply have been irretrievably lost. It has been demonstrated that cash loans will be repaid even in the most difficult circumstances but donors have not yet grasped the point. Although donors provided some funds, they tended to dismiss this kind of activity as peripheral. The World Bank report cites a case in 2000 where an entire programme was undermined while the NGO waited months for approval from the donor who clearly had little idea of the importance of timing (Beck, 2005, p 20). Donors themselves remained focused on infrastructure, allocating what the NGOs considered to be disproportionate funding.

Ideally, all phases of the disaster response should be fully funded, but in practice massive spending by donors on relief and reconstruction reduces the funding available for livelihoods and leaves 'recovery' as the neglected part of the process. One of the main reasons for this may be the institutional divide between 'humanitarian' and 'development' sections within agencies, and the availability of different budgets. Underlying this may also be the tendency to prefer what donors can provide rather than what they can support in terms of local self-help.

An evaluation of the World Bank's general responses to natural disasters finds that the Bank's interventions have been conditioned, to some extent, by its mandate as a development organisation. It has tended, at least in some cases, to ignore disasters as an integral part of the development process and see them as extraneous events: 'Disaster is still sometimes treated as an interruption in development rather than as a risk to development' (IEG, 2006, p ix).

To some extent, the humanitarian system works in a symbiotic manner and the tendencies of donors can be counteracted by the tendencies of NGOs. If donors wish to spend their funds on relief and reconstruction, then public donations can be held back for recovery through NGOs. The system can be self-correcting through market forces. But the scale of resources available to NGOs without donor backing is small compared to what donors can provide. Public appeals are rare events – typically no more than one or two in a year. Even when they happen, there can be strong pressure to spend funds fast, meaning on relief. Despite massive public donations for the Tsunami Disaster, the TEC evaluations found that agencies had not adequately addressed the recovery phase. The problem is deeply embedded in the system, reflecting a tendency to do what is dictated by external pressures and institutional structures rather than what the affected people demand.

Many of these problems arise from drift in the system rather than any conscious intention. WFP has become powerful because there is a lot of money in food aid. The decision to create special rights for refugees was not meant to lead to the neglect of other vulnerable groups. Agencies do not deliberately ignore livelihoods and recovery. But lack of external challenge allows these trends to continue and consolidate. Evaluation sometimes raises questions, but in itself has too little influence to overcome the inertia of the system.

2.4 Local interests

2.4.1 Local political interests

Long-term humanitarian crises generate substantial vested interests both inside and outside the humanitarian sector, and the longer they last the more powerful these interests become. The evaluation of responses to the 2002–3 emergency in Ethiopia predicts that: 'The emergency is not over for 7.8 million people who remain food insecure in 2004 and there is a high probability that more emergencies with increased frequency will occur, perhaps much larger than the 2002–3 emergency' (Steering Committee, 2004). Food aid for Ethiopia has become institutionalised. But interests in this process are not limited only to the food-aid lobby.

Food aid for Ethiopia is an important source of income for trucking operations controlled by the dominant political party. Arguably, international responses to the food crisis affect the political balance of power. Cash distributions give more choice to local people and in a sense encourage a broader form of democracy, whereas food aid has a hierarchical quality, especially when delivered in trucks controlled by a political party.

Shifts in the way aid is delivered may give rise to new interests. A USAID evaluation of cash distributions in Ethiopia finds that a political economy may be developing around this newer approach:

> The cash distribution for beneficiaries gave the tax collectors another opportunity to collect the land tax. The appropriateness of the government collecting taxes from the most vulnerable households who have just received allotments intended for the purchase of food and asset replenishment was an unresolved issue. (Brandstetter, 2004, p 20)

When a government decides whether or not to call for future assistance, factors such as income to the party through trucking and tax collected from cash distributions may be weighed up – but behind closed doors.

Evaluation rarely reaches deep enough to explore such processes. This has more often been left to academic research, which may produce results only years after the event. In his book *The Benefits of Famine*, published in 1994, David Keen showed how politicians and military leaders ruthlessly manipulated food aid during the Sudan war in the 1980s (Keen, 1994). By disrupting supply lines, merchants in collusion with politicians were able to create distress conditions that forced people to sell off cattle at deflated prices. Unfortunately, aid managers at the time were only hazily aware of this (Vaux, 2001, pp 69–92).

There is every reason to believe that such processes continue today and exert an influence, in various subtle ways, on humanitarian responses. Politicians in Sudan today benefit from the lucrative contracts for flights to Darfur and the rise in house rents in Khartoum. They have massive financial interests in the aid operations. These might even weaken their determination to end the war in Darfur. Evaluation tends to stop at the level of formal documentation and avoid such issues. But with increasing volumes of humanitarian aid, interests will increase. The need for parallel processes of research is also becoming greater.

2.4.2   Trends and personal preferences

The humanitarian system inevitably generates its own biases in the form of personal preferences. It suffers from the fads and fancies that pass through the system, creating ripples of distortion. Mostly these do little harm and could even be viewed as evidence of diversity and creativity. But sometimes they interact with deeper forces, and create serious rifts and risks.

Western aid inevitably reflects Western sentiment. The humanitarian system is particularly well geared to address the needs of children, with strong specialised agencies such as UNICEF and Save the Children. To an extent, the focus on children reflects a Western romantic view of innocence. Generous funding may be available for children while there may be suspicion about adults. Child sponsorship is a particularly successful fundraising mechanism but creates problems for aid managers who have to discriminate not only between children and adults but between sponsored and un-sponsored children. Humanitarian principle, notably the Red Cross Code, is clearly against such discrimination and most of the specialised agencies have long ago shifted their focus to children in the community rather than individual children.

The issue of focus on women is more complex, and even more susceptible to the personal interests of aid workers. An increasing proportion of women among aid workers has helped to focus attention on the importance of differentiation by gender in the analysis of needs. Equality for women, in the form of non-discrimination, is strongly endorsed by the Red Cross Code and Sphere, and supported by international law.

Problems arise when aid workers focus exclusively on women and put forward particular views about the status of women in society. The exclusive focus on women has been named the 'Women in Development' approach and was for some years a dominant trend in Western aid, with the spread of projects specifically for women in isolation from other needs and social linkages. There is now wider recognition that such projects should be integrated into a wider process of increasing women's representation.

Several evaluations in the ALNAP dataset make observations about the focus on women. A review of approaches to gender issues in Afghanistan warns that 'Our

experiences reveal that gender in Afghanistan translates, de facto, into women.... Women as a category are singled out in isolation from their wider social, cultural and family contexts' (Abirafeh, 2005, p 13). In the context of a Western invasion and the continuing Global War on Terror, focus on the status of women can easily be perceived as an attack on Islam. The report reflects this perception: 'Western-originated approaches... have more to do with international politics and agendas of external agencies than they do with meeting the felt and expressed needs of the majority of Afghan women' (Barakat and Wardell, 2004, p 19, cited in Abirafeh, 2005).

Another aspect of the tendency to focus on women and girls is the risk of neglecting men and boys. An evaluation of ECHO activity in Sri Lanka and the refugee camps in Tamil Nadu finds that the vulnerability of boys is neglected:

> Returning [male] child soldiers are often severely traumatised and face considerable difficulties in reintegrating into society. In addition they have been recruited while they were still of school-going age. When they return they have missed essential years of schooling and training, and are unable to make up for them. (Herms et al, 2005, p 62)

The same report finds that very little attention is given to these young men, while by contrast attention is given to women and also to child soldiers because 'being a child soldier is also being a victim, and naturally deserving assistance'.

But the issue of boys who are sent away by their parents to avoid recruitment is relatively neglected. They miss out on education and are likely to end up in unemployment. They suffer from lack of self-esteem and this leads to widespread drug-taking. A man who has escaped fighting at the expense of education may then get married: 'to a girl that unlike him has been able to finish her schooling and training, and is therefore earning the family income'. Violence and aggression are common in such situations. The ECHO report also notes that, 'The children, especially the boys, eventually drop out and have a good chance to repeat their fathers' biographies' (Herms et al, 2005).

Girls are also drafted in to the fighting in Sri Lanka, and suffer serious consequences also, but boys are more likely to be sent away. After losing opportunities for education they suffer greater difficulties in performing the gender roles assigned to them, such as being 'head of the family' and 'breadwinner'. In this cultural context,

'a real man has a job, and is not in charge of the household chores' (Herms et al, 2005). Are such boys neglected because Western aid workers do not sympathise with the gender roles that have caused their predicament?

An evaluation of psychosocial responses after the Bam earthquake in Iran (Colliard, 2005, pp 43, 45, 48) also draws attention to the relative neglect of boys. Research showed that boys were more deeply hurt by their experiences than girls. They responded less well to the artistic and handicraft activities that were generally being provided than to sport. The report concludes that further efforts are necessary to understand the needs of boys and to devise responses that are sensitive to these needs.

There are, of course, valid reasons for focusing on women and girls. In most cases, their interests are poorly represented and need to be emphasised. They may be more vulnerable to security threats. But a Western concern for narrowing the gap between gender roles should not be transposed automatically into a focus on women and girls, especially in the current global context in which the West is often suspected of manipulating people and societies for its own ends. Above all, aid workers have to ask themselves whether they are addressing issues of humanitarian concern, or transposing their own concerns and perceptions onto another society. This is not easy.

2.5   Addressing the problems

2.5.1   Donors

During the 1980s it had been necessary to label humanitarian aid as impartial or neutral in order to gain access to areas under communist influence and to avoid potential confrontation. After the end of the Cold War humanitarian aid in Africa and Asia came under greater pressure from political and security interests. In the 1990s the UK government, for example, developed a policy of 'coherent' responses in which aid, foreign policy and military considerations were joined together in a

single strategy. This formed the basis for the 'Global Conflict Prevention Pool' in which the Ministry of Defence, the Foreign Office and the Department for International Development (DFID) share pooled resources and are encouraged to develop joint approaches (Harmer et al, 2004). A report in the ALNAP dataset on Finnish aid gives a good indication of the position of humanitarian aid in Western governance. It quotes official documents, asserting that 'humanitarian assistance… should be seen as an integral part of Finland's foreign policy, with definite linkages to the country's security policy, development aid policy and human rights policy' (Telford et al, 2005, p 14).

Both DFID and Finland are prominent members of the Good Humanitarian Donorship (GHD) initiative which, launched in June 2003, is intended to encourage a more collaborative approach among donors and a tighter focus on needs. It has adopted the principle of impartiality, defined as 'the implementation of actions solely on the basis of need, without any discrimination between or within affected populations.'[14]

The GHD initiative has been piloted in the DRC, where a set of indicators was developed to measure the application of the GHD principles, and it has also been evaluated in Burundi (Bijojote and Bugnion, 2004). The general conclusion is that progress has been distinctly limited. The Burundi evaluation finds that the objectives remained poorly understood, and there was very limited buy-in from the stakeholders. The evaluation of ECHO-financed programmes in Burundi comes to a similar conclusion. The GHD initiative has produced 'little discernible result so far in Burundi…. Donors appear to be too often designing their strategies in isolation of one another, which is leading to fragmentation and widely divergent application of policies' (Shepherd-Barron and Fadiga, 2005).

In practice the GHD initiative seems to represent good intentions rather than any actual change in donor practice. As the UN Humanitarian Response Review concludes:

> The uneven support given to forgotten emergencies or neglected needs present the humanitarian organizations and the donors with the dilemma of the equity of their response. The steps taken in the framework of the GHD initiative are an attempt to address the dilemma but with an attitude, on the side of the donors, which remains too timid. (Adinolfi et al, 2005, p 59)

2.5.2   The United Nations

As noted above, the UN consists of a number of ad hoc structures that do not fit well together and often act as interest groups rather than a single structure. Accordingly the issue of coordination has been particularly thorny. The UN Humanitarian Response Review finds that coordination is poor between the three main levels of implementing agencies in the humanitarian system – UN, Red Cross movement and NGOs (Adinolfi et al, 2005, pp 10–11). This may reflect the trend among donors towards funding through bilateral rather than multilateral channels, and particularly through NGOs. The Tsunami Disaster showed NGOs for the first time in a position to dominate the UN, but unable to organise themselves into an effective coordination system (Telford and Cosgrave, 2006, Bennett et al, 2006).

The IDP synthesis report indicates that relations between donors and the UN have not improved. The donors 'are not doing nearly enough to support coordination mechanisms whether for overall humanitarian efforts or those specifically relating to IDPs' (Borton et al, 2005, p 17). Overall the trend seems to be towards more vigorous action by donors in pursuit of their security interests. This allows them to collaborate with each other but does little to encourage collaboration throughout the system. Since the UN is already a deeply fractured structure in relation to humanitarian action, it seems likely that opportunism rather than principle will prevail.

As the UN Humanitarian Response Review concludes:

> Humanitarian organizations and donors acknowledge that the humanitarian response provided is not good enough, and that remedial action is needed and a number of initiatives are currently being taken to address this. Such initiatives focus on accountability to direct beneficiaries, donors and taxpayers, national or local authorities, as well as governing bodies of the humanitarian organizations. The major challenges are to reconcile different, if somewhat contradictory imperatives, to define the appropriate limits of accountability and to ensure that the accountability agenda is driven by the humanitarian principles and the needs of the beneficiaries. (Adinolfi et al, 2005, p 9)

2.5.3 Non-governmental organisations

From 2000 to 2004 NGOs produced a number of reports on the issue of 'forgotten emergencies' (Oxfam, 2000; 2003; Cosgrave, 2004; Christian Aid, 2004) but these were not followed up with campaigns, and more recently the NGOs have been silent on this topic. It has been argued that NGOs have now become fully integrated into a donor-driven humanitarian system that acts primarily in the interests of Western security (Duffield and Waddell, 2006). In such a system, distortions in relation to need are inevitable.

NGOs have also been criticised, notably in the TEC synthesis report (Telford and Cosgrave, 2006) for lack of engagement with the Western public over this issue. With some exceptions, they have made only half-hearted attempts to challenge the tendency to give money in response to images in the media rather than by considering a wider range of humanitarian need. NGOs could be accused of exploiting the ignorance of the Western public rather than making concerted efforts to bring about a better understanding and more rational choices.

Oxfam's Peer Review after the Bangladesh floods indicates that NGOs rely on internal debate in order to curb such extremes (Oxfam Evaluation Team, 2005, p 22). Agencies that question themselves and learn from their experiences continue to move forward. But they are under no compunction to do this. The humanitarian system continues to demonstrate extremes of good and bad practice, and it is by no means clear that market forces favour good practice. This has led some commentators, including the authors of the TEC synthesis report, to call for more regulation.

A striking feature of the ALNAP dataset is that NGOs are clearly unwilling to publish their evaluations. The Sphere Common Standard on evaluation requires that: 'There is a systematic and impartial examination of humanitarian action, intended to draw lessons to improve practice and policy and to enhance accountability' (Sphere Project, 2004 p39).

Accountability, whether to donors or beneficiaries, must depend on transparency, which, in the case of evaluation, means publication.

2.5.4   The problem of needs

As this review has shown, the humanitarian system is driven by supply rather than demand. But there remains a further problem of defining what 'demand' really refers to. It cannot be simply a list of 'wants' but must refer to basic human requirements. But this is also hard to define. Current problems stem from the tendency to focus on immediate physical requirements as 'needs' and to demote vulnerability to a remote contingency. In reality, vulnerability refers to immediate threats to security and the collapse of the means of livelihood following a disaster. Agencies cannot retrospectively address the vulnerability that preceded a disaster, but they can focus on the vulnerability that follows.

In the Tsunami response, huge amounts of money were given to survivors but after the first few days there was no real threat to human life. People wanted to rebuild their lives or 'build back better' but they were not immediately vulnerable. The tragedy of the DRC was that its people remained vulnerable throughout the response, and therefore money would have saved lives. It follows that the focus of humanitarian action should be vulnerability, rather than need. Or, as Darcy and Hofmann conclude:

> The concept of need as deficit, and consequent deficit-based analysis, reinforces the tendency to define need in terms of goods and services on offer, which people are found to lack. Instead of an analysis based on the ambiguous concept of need, the study recommends one based on acute risk, understood as the product of actual or imminent threats and vulnerabilities. (Darcy and Hofmann, 2003, p 5)

There are now serious attempts to reform the Consolidated Appeals Process. The UN is developing an index of humanitarian need. ECHO has already adopted the language of vulnerability and is using a Vulnerability Index to decide on the allocation of aid (ECHO, 2006). But it remains to be seen whether the humanitarian system will be able to agree on a common method of 'needs assessment'.

2.5.5   Accountability to whom?

Another theme emerging from this review of evaluations is a persistent uncertainty about the relationship between 'accountability to donors' and 'accountability to beneficiaries'. The Red Cross Code sidesteps the issue by putting both together, without acknowledging that the two forms of accountability can pull in different directions. Throughout this chapter there have been examples of conflicting forms of accountability, such as the contrasts between supply-driven and demand-driven responses. Evaluators often question 'supply-' or donor-driven responses and argue for greater priority to be given to 'demand-' or beneficiary-driven responses. In the evaluations comprising the ALNAP dataset, agencies are often urged to give greater attention to consultation and participation.

An emphasis on vulnerability could be integrated with a priority for demand-driven responses through the notion of 'human security'. This was introduced in the UNDP *Human Development Report* of 1994 and further developed by the Commission on Human Security which defined human security, following Amartya Sen, as 'freedom from fear' and 'freedom from want' (Commission on Human Security, 2003, pp 8–9). Sen deliberately framed human security in terms of what people perceive, and emphasised the potential for 'downside risk', corresponding to 'vulnerability'. The concern of outsiders should be that people do not lose what they already have, however meagre.

The concept of human security has the merit of uniting relief, development and conflict perspectives, and embracing both need and vulnerability. But it has the disadvantage that its subjective form makes it difficult to put into practice. The Commission on Human Security has translated 'freedom from fear' into 'protection' and 'freedom from want' into 'empowerment', but these are inexact equivalents.

Unfortunately the notion of human security has already been co-opted as a way of extending the language of global security to the local or 'human' level (Duffield and Waddell, 2006). But this does not mean that it can no longer be used in its original form. An index of human security might be preferable to an index of humanitarian 'need'. But it may be even more important to find common ground.

2.6   Conclusions and ways forward

2.6.1   Summary of conclusions

Humanitarian aid may be the main focus for aid workers but it is not the centre of attention for most of the world. Political interest and public sentiment play a massive role in the allocation of aid. Aid is not directed where it is most needed. Where humanitarian action achieves a high profile, it is particularly susceptible to pressures that reduce the quality of the response. Especially in cases of conflict, humanitarian actors have not equipped themselves with the tools to analyse the context adequately, and remain susceptible to the political interests of donors and other parties.

Over the years the humanitarian system has developed structures that reflect sentiment and interest rather than need. It has also built an architecture of agencies that, especially in the case of the UN, divides problems and people in unsuitable ways. This not only confuses the response but also distorts it, notably towards food aid and refugees, and creates an inertia that makes change and adaptation difficult. The system is better at addressing the needs that it can fulfil than it is at assessing vulnerability. Consequently, it is poor at prevention and devotes most of its energy to problems that are already past. It draws staff from all over the world to address the aftermath of a Tsunami Disaster, when there are few lives to save, but neglects unnecessary deaths in their millions in DRC.

In disaster response, the humanitarian system still tends to favour relief and to neglect recovery, especially in relation to livelihoods. Donors in particular tend to jump from relief to physical reconstruction without supporting affected people in their attempts to earn their own living. This reflects a lack of responsiveness in the system to the demands of the affected people. Only rarely is their voice heard, often through local NGOs, but this is transposed into action only if they are powerful enough to wield an influence within the humanitarian system.

For the Red Cross movement and international NGOs, the Red Cross Code is the 'industry standard' but is being breached on a massive scale. Principle Four enjoins agencies not to act as instruments of government foreign policy. But since foreign policy is a fundamental element in the global allocation of aid, and NGOs depend heavily on institutional funding, humanitarian action follows politics. Principle One

of the Red Cross Code requires agencies to direct their attention to need above all other considerations, and this is endorsed by the Sphere Charter and Common Standards. But NGO responses are heavily skewed by public appeals. The pattern of humanitarian aid by NGOs across the world is much the same as for donors.

The consequences of imbalances in humanitarian aid are significant. In situations of high profile, such as the Tsunami Disaster, agencies become preoccupied with speed and focus on token successes rather than real needs. Staff and management capacity is drawn from other areas, and local capacity is bypassed. In the most neglected cases, such as in DRC, the result is unnecessary deaths, but to a limited extent agencies compensate by a closer relationship with local communities, leading to what may arguably be more sustainable solutions. In countries of dominant donor interest, such as Iraq, international NGOs have set a poor example to local NGOs, especially in relation to independence. As a result, NGOs and the Red Cross movement are increasingly perceived as a tool for the extension of Western interests.

At the local level, aid actors also tend to see themselves at the centre of the action, when in reality they may be tools of local interest. Sometimes they pursue their own interests and preferences, muddling rights with needs and overlooking vulnerable groups in the process.

Needs assessment remains the fundamental flaw of the humanitarian system. There is no accepted method of assessment. Comparisons are difficult and distortions become apparent only when they are absolutely glaring, as in the case of the neglect of DRC. UN Appeals are often used by donors as the basis for aid allocations but bear little relationship to need. Indeed, 'need' is often reduced to little more than a list of what the UN agencies can most easily provide. Vulnerability, particularly in the form of lack of protection against threats to security, has been seriously neglected. Vulnerability is a better basis for assessment than need, but the system is only just beginning to define this term and find ways to measure and respond to it. More comprehensive notions, such as human security, have yet to be developed and applied.

Efforts to address these issues are tentative and inadequate. Donors joining together in the Good Humanitarian Donorship initiative have endorsed the principle of impartiality, and yet global allocations of aid continue to be seriously distorted and the initiative has had little impact on the ground. The 'Global War on Terror' has

introduced a new focus on global security that has trumped all other considerations. It remains to be seen whether donors can hold the line against such pressures. International NGOs and the Red Cross movement have yet to make any serious challenge to the distorted system in which they are embedded.

Structural problems will not be solved by collaboration but may need more vigorous forms of coordination. The problem is not specifically in the relationship between NGOs and donors, or between the UN and the Red Cross movement. It lies in the vigour of the top-down dynamic, driven by political and media interests. The answer is to strengthen bottom-up processes that provide better representation for beneficiaries. This can be done by vigorous assertion of humanitarian principle through designated leadership.

2.6.2   Ways forward for donors and the UN

The ALNAP dataset for 2005 offers an opportunity to examine humanitarian practice during the period when the Global War on Terror has reached its maturity and is seen competing with more classical humanitarian perspectives. The humanitarian system is shown to be swayed almost off its moorings by the influence not only of the global security agenda, as in Iraq and Afghanistan, but also by poorly informed and imbalanced responses by the Western public, as in the case of the Tsunami Disaster. There is no sign that such massive pressures are likely to abate. The humanitarian system is not strong enough to challenge the global security agenda. NGOs have shown little inclination to challenge the Western public. The immediate problem is how to minimise the damage.

The humanitarian system will be unable to cope with external threats so long as it remains divided and competitive. Efforts to promote more collaboration, such as the Good Humanitarian Donorship initiative, must be applauded and supported. But the limited results so far suggest that this process will not go far enough to make a significant difference. Lack of impact on the ground indicates that the process is 'top down'.

The West's increasing willingness to intervene in other countries has led to a readiness to 'manage' conflict though aid interventions. In the case of Afghanistan, strategies for humanitarian aid are overtly developed through a security perspective.

Some elements within such a strategy will be valid in terms of humanitarian response, while others may be invalid. Disentangling the elements requires systematic analysis of the nature of the security threats and their roots in issues of humanitarian concern. Aid actors need their own forms of conflict analysis if they are to maintain their independence.

Further efforts must be made to counteract distortions in the global allocation of aid. If the UN Appeals are to be used as the basis for such allocations, they must be made systematic and objective. Work by the UN and ECHO to establish indices of humanitarian need should be examined carefully in order to reach a common agreed standard which can then be used as the prime basis for aid allocations, separating out the political and security interests. For the best representation of the security interests of 'beneficiaries', a human-security approach deserves consideration. Only when better assessment of need is available can there be a basis for accountability.

In disaster response, donors should focus more attention on the transition or recovery phase. This tends to be neglected for institutional reasons. In particular, recovery through livelihoods needs a higher level of funding and should be recognised at an early stage.

It may be unrealistic to suggest that the UN should dismantle its agencies and start all over again, but firm central control is needed in order to tackle the 'box' mentality that arises from such a system, and its inevitable tendency to follow vested interests. The limited success of the 'collaborative approach' of the UN in relation to IDPs indicates that collaboration without authority will lead to weakness, and this will leave the UN open to 'distorting' influences. A more vigorous style of leadership based on representation of humanitarian interests might be better for the system as a whole.

2.6.3   Ways forward for NGOs and the Red Cross movement

Codes and principles are an essential defence against distortion. The Red Cross Code and Sphere Common Standards are the nearest thing available to a statement of rights for beneficiaries. Sphere acknowledges that, although there are many different forms of accountability, 'our fundamental accountability must be to those we seek to assist' (Sphere Project, 2004, p 19). In current times this needs to be

actively asserted. Similarly, the first Red Cross Code principle, that need comes first, should be actively promoted. Agencies should tackle the distortions of the global humanitarian system in which they are embedded.

Box 2.3	Beneficiary surveys

In balancing 'accountability to donors' with 'accountability to beneficiaries', a key problem is finding out what beneficiaries think. Although public-opinion surveys are a basic tool in many other sectors, they are rare and under-developed in the humanitarian sector. One of the first evaluations to use large-scale beneficiary surveys was the DEC evaluation of the response to the Gujarat earthquake. A survey covering around 2,300 people conducted by a local organisation, the Disaster Mitigation Institute (DMI), found that beneficiaries were more critical of the response than had been expected by the general evaluation team.

The process was repeated in the DEC evaluation of the response to the Tsunami Disaster in 2005, with surveys in India, Sri Lanka and Indonesia using local respondents. The comments from beneficiaries were again found to be quite critical and provided a useful check on the evaluation as a whole, which had otherwise relied heavily on inputs from the agencies themselves. In particular, the survey showed concern that livelihoods had been neglected and that the transition to recovery had been too slow. Similar surveys, with similar results, were conducted as part of the TEC Capacities and LRRD evaluations (Scheper et al, 2006), Christoplos, 2006). These surveys, nine months after the disaster, contrasted with surveys conducted by the Fritz Institute in May, just five months after the disaster, and helped to show that the transition from relief to recovery had not been properly handled. This suggests that beneficiary surveys are ideally conducted on a longitudinal basis, to map changes in public opinion over time.

The methodology for such surveys still needs to be developed. The DMI team has found that anecdotal information can be used in a creative way while remaining systematic. Cross-checking, triangulation and questioning within the survey team are important elements in the process. Surveys are an important step towards 'accountability to beneficiaries' but it is a poor reflection on the humanitarian system that they are so rarely used.

This should ideally be done jointly, but is this realistic? Although coordinating bodies have long existed, they have not been prominent in tackling major threats such as the Global War on Terror and distortion of the humanitarian system. The competitive behaviour that characterised the response to the Tsunami Disaster is not an encouraging sign. Freed of UN coordination, NGOs showed little inclination to coordinate or place the common good above their own interest. But perhaps the criticisms they have received from the TEC evaluations and elsewhere are just the jolt that is needed to bring about new forms of collaboration.

International NGOs and the Red Cross movement also faced strong criticism from the TEC evaluations for failing to establish adequate mechanisms for consulting, informing and involving the people affected by humanitarian disasters. With donors and the UN now rethinking the definition of 'need', these agencies have an important role to fulfil in representing beneficiary views. What really should be the concern of outsiders – need, vulnerability or human security? Perhaps a lead can be taken from beneficiary surveys (Box 2.3). These have provided important new perspectives on humanitarian responses and could be used to help form this definition. Once a global definition of need is available, NGOs should be able to hold donors to account, and strive to reverse the current polarity of power.

2.6.4 Summary of recommendations

Recommendations to donors and the UN

- Reinvigorate the Good Humanitarian Donorship initiative, focusing on the issue of impartiality in global aid allocations and the challenge of establishing methods for measuring and comparing needs.

- Initiate processes of conflict analysis, especially where there may be a significant difference between approaches based on global security, state security and human security.

- Address the deficiencies of the Consolidated Appeals by establishing indices of need that take into account vulnerability and human security.

- Focus greater attention on the recovery phase of disaster response.

- Support collaborative approaches in the UN but do not ignore the need for dynamic leadership based on humanitarian principle.

Recommendations to NGOs and the Red Cross movement

- Use the principles of the Red Cross Code and Sphere to uphold humanitarian values in the face of distortions caused by political and media pressures.

- Engage with other actors to ensure that 'need comes first' can be translated into a way of measuring and comparing aid allocations.

- Develop mechanisms to ensure that 'accountability to beneficiaries' takes precedence over 'accountability to donors'.

- Make use of beneficiary surveys to support and guide the activity outlined above.

- Make use of formal research in order to confront systemic failures in the humanitarian system.

- Engage with the Western public and media to promote better understanding of comparative need.

- Publish more evaluations.

- Promote internal accountability and debate through peer review.

- Focus on public audit as the primary method of downward accountability, and promote this method among donors and UN agencies.

Notes

The author was directly involved in some of the reports referred to in this chapter, including the DEC evaluations of responses to the Gujarat earthquake and the Indian Ocean Tsunami. He contributed to the TEC report on the impact of the international tsunami response on national and local capacities.

1 IFRC (1994): www.ifrc.org.

2 'NGOs' here includes the Red Cross, although there are no Red Cross reports in the dataset.

3 Public Service Agreement, p 19 – see DFID website: www.dfid.gov.uk.

4 For a detailed comparison of countries receiving more and less aid, see Development Initiatives (2005), p 22.

5 As an example: 'In Faridpur, sanitary napkins were not included in the package on the grounds that the donor did not allow for it. This gap could have been easily filled with Oxfam's own resources or we could have lobbied with the donor accordingly' (Oxfam Evaluation Team, 2005, p 15).

6 Personal communications.

7 International Rescue Committee survey data updated for *The Lancet*, 7 January 2006. See www.theirc.org/index.cfm/wwwID/2129.

8 Estimate by the author based on data from Development Initiatives (2005).

9 Total humanitarian assistance over the main period of insecurity, 1995–2004, was $265m. Bosnia-Herzegovina received $1,644m and Serbia/Kosovo $1,299m (Development Initiatives, 2003, pp 32–33).

10 'A factor that has limited funding is the finite number of willing and able implementing partners and, in the past, limited humanitarian space' (Michael et al, 2004, p 20).

11 The only example is Currion (2005) which focuses on information technology and describes itself as an 'assessment'.

12 DFID has identified improvement of the CAP process as a target in its Public Service Agreement. It is also a focus for the Good Humanitarian Donorship initiative.

13 Harvey (2005) provides a useful summary of the literature. See below for more on this.

14 See Harmer et al (2004) for a review of objectives and progress.

Evaluation dataset for the synthesis

Abirafeh, L (2005) *Lessons from Gender-focused International Aid in Post-conflict Afghanistan…. Learned?* Friedrich Ebert Stiftung.

Adinolfi, C et al (2005) *Humanitarian Response Review.* UN. Team members: David Bassiouni, Halvor Lauritzen and Howard Williams.

Bailey, J et al (2005) *Full Report of the Thematic Review of Targeting in WFP Relief Operations.* World Food Programme. Team members: Jeremy Shoham, Kay Sharp, Egon Westendorf, Jeffrey Marzili, Helen Young, Ben Watkins

Beck, T (2005) *Learning Lessons from Disaster Recovery: The Case of Bangladesh.* Disaster Risk Management Working Paper Series, No 11, World Bank.

Bijojote, S and C Bugnion (2004) *External Baseline Evaluation of the Burundi Good Humanitarian Donorship Pilot.* UNOCHA.

Borton, J et al (2005) *Support to Internally Displaced Persons - Learning from Evaluations, Synthesis Report of a Joint Evaluation Programme.* SIDA.

Brandstetter, R (2004) *Evaluation of OFDA Cash for Relief Intervention in Ethiopia.* Checchi and Company Consulting/USAID

Brennan, M et al (2005) *Refugee Health in Nepal - Joint UNHCR-WHO Evaluation of Health and Health Programmes in Bhutanese Refugee Camps in Nepal.* UNHCR. Team members: Oleg Bilukha, Marleen Bosmans, BR Dahal and Kandarpa Chandra Jha.

Carloni, A et al (2004) *Summary Report of the Evaluation of Afghanistan PRRO 10233.* World Food Programme. Team members: Bernard Broughton, Peter Sloane, Silvia Kaufman, Fazalkarim Njimi, Jonas Lindholm, Pernille Hougesen and Salem Shah.

Carlson, M et al (2006) *WFP-UNHCR Joint Evaluation of the Pilot Food Distribution Projects – Volume I and II.* World Food Programme. Team members: François de Meulder, Paolo Mattei and Julian Lefevre.

Colliard, C (2005) *A Psychosocial Programme of Recreational Centres in Bam (Iran).* Terre des Hommes/Centre for Humanitarian Psychology.

Creti, P (2005) *Evaluation of the Livelihood Programmes in Mapou and Cape Haitian, Haiti.* Oxfam GB.

Currion, P (2005) *Assessment Report: Pakistan Earthquake Response November–December 2005.* Emergency Capacity Building Project: Information Technology and Requirements.

Dammers, C et al (2005) *Report for the Evaluation of ECHO's Actions in the Democratic People's Republic of Korea 2001–4.* AguaConsult/ECHO. Team members: Patrick Fox and Michelle Jimenez.

Danida Evaluation Department (2005) *Humanitarian Assistance to Afghanistan, 2001–5.* Danida. Consultants: Christian Michelsen Institute, Norway (lead), Copenhagen Development Consulting and the German Association of Development Consultants.

Durán Vargas, LR and J Kressin (2004) *The Evaluation of DIPECHO Action Plans in the Caribbean Region.* German Association of Consultants/ECHO.

Etherington, A and E Roberts (2005) *Report of the External Evaluation of the ECHO funded Save the Children DPRK Programme.* ECHO/SCF.

Goodhand, J (2005) *Aid, Conflict and Peacebuilding in Sri Lanka 2000-2005.* Sri Lanka Strategic Conflict Assessment, Vol 1, Asia Foundation. Team members: Bart Klem, Dilrukshi Fonseka, SI Keethaponcalan and Shonali Sardesai.

Gregg, B (2004) *Evaluation of OFDA Emergency Seed Relief Drought Response, Ethiopia 2003–4.* USAID.

Groves, L (2005) *UNHCR's Age and Gender Mainstreaming Pilot Project 2004.* UNHCR.

Hansen, G (2004) *Independent Evaluation: Iraq NGO Coordination and Security Office (ECHO/IRQ/210/2003/05029).* NCCI.

Hassan, U et al (2005) *Evaluation of the Bangladesh Emergency Flood Response.* Oxfam. Team members: S Ireland and N Kukrety.

Herms, E-M et al (2005) *The Evaluation of ECHO's Actions in Sri Lanka and in the Tamil Refugee Camps in Tamil Nadu, India.* German Association of Conultants/ECHO. Team members: Richard Ellert, Alice de Jonge and Leonie M Barnes.

IEG (Independent Evaluation Group, World Bank) (2006) *Hazards of Nature, Risks to Development: An IEG Evaluation of World Bank Assistance for Natural Disasters.* World Bank.

Landart, G et al (2005) *Evaluation of WFP's Assistance to China (1979–2005).* World Food Programme. Team members: Sylvie Dideron, Henri Loze, Cui Shuli, Pernille Hougesen and Celso Murano del Picchia.

Michael, M (2004) *Evaluation of GOAL's ECHO-funded Health Intervention in North Katanga, Democratic Republic of Congo.* GOAL.

Michael, M et al (2004) *Evaluation of ECHO's Actions in the Democratic Republic of Congo (DRC).* ECHO. Team members: Véronique Thouvenot, Thomas Hoerz and Laura Rossi.

Michael, M et al (2005) *The Road to Health and the Road to Afghanistan: Joint UNHCR/WHO Evaluation of Health and Health Programmes for Afghan Refugees Living in Pakistan.* Other team members: Mary Corbett and Glen Mola.

Napier, A (2004) *The Story of RBA in CARE Sierra Leone: Demonstrating Why a 'Rights-based Approach' is a Better Way of Conducting Humanitarian Endeavors.* CARE International.

Oxfam (2006) *Evaluation of Livelihood Components of Oxfam's Protracted Relief Programme in Zimbabwe to date, and Communities' Needs Identification.* Target Research.

Pougin de la Maisonneuve, C et al (2005) *Refugee Health in Zambia: Joint UNHCR–WHO Evaluation of Health and Health Programmes in Refugee Camps in Zambia.* UNHCR. Team members: Manengu Musambo and Roselidah Ondeko.

Rijsdijk, A and M Andreini (2005) *Evaluation of ECHO's Financed Actions in Haiti.* GFE/ECHO.

Shepherd-Barron, J and F Fadiga (2005) *Evaluation of ECHO's Financed Actions in Burundi.* GFE/ECHO.

Steering Committee for the Evaluation of the Joint Government and Humanitarian Partners Response (2004) *Evaluation of the Response to the 2002–03 Emergency in Ethiopia.*

Telford, J and J Cosgrave (2006) *Joint Evaluation of the International Response to the Indian Ocean Tsunami: Synthesis Report.* London: TEC. Contributing author: Rachel Houghton.

Telford, J et al (2005) *The Evaluation of Finnish Humanitarian Assistance 1996–2004.* Team members: Elisa Mikkola, Patricia Bianchi and Arno Kourala.

UNICEF Evaluation Office (2004a) *Assessment of UNICEF's Contribution to UN Reform and its Impact on UNICEF: UN Reform under the UN Development Group.* UNICEF.

UNICEF Evaluation Office (2004b) *Evaluation of UNICEF Emergency Preparedness and Early Response in Iraq.* UNICEF.

Vargas, LRD and J Kressin (2004) *The Evaluation of DIPECHO Action Plans in the Caribbean Region.* Brussels: AGEG/ECHO.

Wilding, J et al (2005a) *Independent Evaluation of CARE's Humanitarian Response to Flooding Resulting from Tropical Storm Jeanne in Haiti (North-west and Artibonite Provinces)*. CARE. Team members: Jacqueline Wood and Yves-Laurent Regis.

Wilding, J et al (2005b) *Joint Independent Evaluation of the Humanitarian Response of CARE, CRS, Save the Children and World Vision to the 2005 Food Crisis in the Republic of Niger*. CARE, CRS, Save the Children, World Vision. Team members: Issaka Idrissa Mossi, Debra Lynne Edwards, Amanda Weisbaum, Boube Aw and Timothy Mander.

Wiles, P et al (2004) *Review of Aid Coordination for Somalia*. SACB. Team members: Khalif Hassan Farah and Adar Abdisalaam Bakard.

Wiles, P et al (2005) *Learning Lessons from Disaster Recovery: The Case of Mozambique*. Disaster Risk Management Working Paper Series No 12, World Bank. Team members: Kerry Selvester and Lourdes Fidalgo.

Note

Some evaluations are listed by commissioning agency or by name of team leader, rather than including the full list of contributors, which may be extensive. In the lists above and below, names of team members are given where possible at the ends of the references.

References and bibliography

ALNAP (2003) *Review of Humanitarian Action 2003*. London: ALNAP/ODI.

ALNAP (2004) *Review of Humanitarian Action 2004*. London: ALNAP/ODI.

Barakat, S (2004) (ed) *Reconstructing War-Torn Societies: Afghanistan*. Hampshire: Palgrave Macmillan.

Barakat, S and G Wardell (2004) 'Exploited by Whom? An Alternative Perspective on Humanitarian Assistance to Afghan Women' in Barakat, S (2004) (ed) *Reconstructing War-Torn Societies: Afghanistan*. Hampshire: Palgrave Macmillan.

Barrett, C and D Maxwell (2005) *Food Aid after Fifty Years: Recasting its Role*. London: Routledge.

Beck, T (2006) *Evaluating Humanitarian Action: Using the OECD–DAC Criteria. An ALNAP Guide for Humanitarian Agencies*. London: ALNAP.

Bennett, J et al (2006) *Coordination of International Assistance in Tsunami-affected Countries*. London: Tsunami Evaluation Coalition.

Berdal, M and D Malone (eds) (2000) *Greed and Grievance: Economic Agendas in Civil Wars*. Colorado and Ottawa: Rienner/IDRC.

Breier, H (2005) *Joint Evaluations: Recent Experiences, Lessons Learnt and Options for the Future*. OECD/DAC Evaluation and Aid Effectiveness Series, No 4, Paris: OECD.

CARE International, Catholic Relief Services, Oxfam GB and World Vision International (2005) *Report of Workshop Proceedings, April 7–8 2005, Bangkok, Thailand*.

Causton, A and G Saunders (2006) 'Responding to Shelter Needs in Post-earthquake Pakistan: A Self-help

Approach' in *Humanitarian Exchange* No 34, June, pp 10–25. London: HPN/ODI.

Chambers, R (1997) *Whose Reality Counts? Putting the First Last.* London: ITDG.

Christian Aid (2004) *The Politics of Poverty: Aid in the New Cold War.* London: Christian Aid.

Christoplos, I (2006) *Links Between Relief, Rehabilitation and Development in the Tsunami Response.* London: Tsunami Evaluation Coalition.

Commission on Human Security (2003) *Human Security Now: Protecting and Empowering People.* New York: UN.

Cosgrave, J (2004) *The Impact of the War on Terror on Aid Flows.* London: ActionAid.

Darcy, J and C-A Hofmann (2003) *According to Need? Needs Assessment and Decision-making in the Humanitarian Sector.* HPG Report 15, London: ODI.

DEC (2001) *Independent Evaluation: The DEC Response to the Earthquake in Gujarat.* London: DEC. Team members: Tony Vaux, Mihir Bhatt, Hugh Goyder, Alex Jacobs and Sarah Routley.

DEC (2005) *Independent Evaluation of the DEC Tsunami Crisis Response.* London: DEC. Team members: Tony Vaux, Mihir Bhatt, Abhijit Bhattacharjee, Michele Lipner, Jean McCluskey, Asmita Naik and Frances Stevenson.

Development Initiatives (2003) *Global Humanitarian Assistance 2003.* Somerset: Development Initiatives.

Development Initiatives (2005) *Global Humanitarian Assistance Update 2004–5.* Somerset: Development Initiatives.

Duffield, Mark and Nicholas Waddell (2006) 'Securing Humans in a Dangerous World' in *International Politics* No 43, pp 1–23. London : Palgrave Macmillan.

ECHO (2006) *Methodological Note – DG ECHO Strategy 2006: Global Humanitarian Needs and Forgotten Crisis Assessment.* Brussels: ECHO.

Harmer, A, L Cotterrell and A Stoddard (2004) *From Stockholm to Ottawa: A Progress Review of the Good Humanitarian Donorship Initiative,* HPG Research Briefing Number 18, October. London: ODI.

Harvey, P (2005) *Cash and Vouchers in Emergencies,* HPG Discussion Paper. London: ODI.

Herson, M (2005) 'Real-time Evaluations in Darfur: Some Suggestions for Learning' in *Humanitarian Exchange* No 30, June. London: HPN/ODI.

Herson, M and J Mitchell (2005) 'Real Time Evaluation: Where does its Value Lie?' in *Humanitarian Exchange* No 32, December. London: HPN/ODI.

IFRC (2005) *World Disasters Report.* Geneva: International Federation of Red Cross and Red Crescent Societies.

IRC (International Rescue Committee) (2003) *Mortality in the Democratic Republic of Congo: Results from a Nationwide Survey.* New York: IRC.

Keen, D (1994) *The Benefits of Famine.* New Jersey: Princeton University Press.

Lawry-White, S (2004) *Evaluation of UNICEF Emergency Preparedness and Early Response in Iraq.* New York: UNICEF.

Macrae, J et al (2002) *Uncertain Power: The Changing Role of Official Donors in Humanitarian Action,* HPG Report 12. London: ODI.

Mendelsohn, S and A Mackenzie (2004) *Assessment of UNICEF's Contribution to UN Reform and its Impact on UNICEF: UN Reform under the UN Development Group.* New York: UNICEF.

Michael, M et al (2004) *Evaluation of ECHO's Actions in the Democratic Republic of Congo (DRC).* Brussels: ECHO. Other team members: Veronique Thouvenot, Thomas Hoerz and Laura Rossi.

Napier, A (2004) *No Rights, No Justice, More War: The Story of RBA in CARE Sierra Leone.* CARE.

OECD/DAC (1991) *DAC Principles for Evaluation of Development Assistance.* Paris: OECD.

OECD/DAC (1998) *Review of the DAC Principles for Evaluation of Development Assistance.* Paris: OECD.

OECD/DAC (2000a) *DAC Criteria for Evaluating Development Assistance.* Paris: OECD.

OECD/DAC (2000b) *Effective Practices in Conducting a Joint Multi-Donor Evaluation.* Paris: OECD.

OECD/DAC (2006) *Guidance for Managing Joint Evaluations*, DAC Network on Development Evaluation. Paris: OECD.

Oxfam (2000) *An End to Forgotten Emergencies?* Oxfam Briefing Paper, May. Oxford: Oxfam GB.

Oxfam (2003) *Beyond the Headlines: An Agenda for Action to Protect Civilians in Neglected Conflict.* Oxford: Oxfam GB.

Oxfam Evaluation Team (2005) *Evaluation of the Bangladesh Emergency Flood Response.* Oxford: Oxfam GB.

Scheper, E, A Parakrama and A Patel (2006) *Impact of the Tsunami Response on Local and National Capacities.* London: Tsunami Evaluation Coalition.

Sen, A and J Dreze (1981) *Poverty and Famines: An Essay on Entitlement and Deprivation.* Oxford: OUP.

Slim, H and A Bonwick (2005) *Protection: An ALNAP Guide for Humanitarian Agencies.* London: ALNAP/ODI.

Smillie, I and L Minear (2004) *The Charity of Nations: Humanitarian Action in a Calculating World.* Connecticut: Kumarian Press Inc.

Sphere Project (2004) *Humanitarian Charter and Minimum Standards in Disaster Response.* Oxford: Oxfam Publications.

Stoddard, A (2003) 'Humanitarian NGOs: Challenges and Trends' in *Humanitarian Action and the Global War on Terror: A Review of Trends and Issues*, HPG Report 14. London: ODI.

Terry, F (2002) *Condemned to Repeat? The Paradox of Humanitarian Action.* Ithaca and London: Cornell.

Vaux, T (2001) *The Selfish Altruist: Relief Work in Famine and War.* London: Earthscan.

Vaux, T (2006) 'Humanitarian Trends and Dilemmas' in *Development in Practice* Vol 16, Nos 3 & 4, June. Oxford: Oxfam GB.

Vaux, T, M Bhatt, A Bhattacharjee, M Lipner, J McCluskey et al (2005) *Independent Evaluation of the DEC Tsunami Crisis Response: Final Report: November 2005.* London: DEC.

Wiles, P and L Sida (2006) *The 2004 Indian Ocean Tsunami Disaster: Evaluation of UNICEF's Response: Synthesis Report.* New York: UNICEF.

Wiles, P et al (2005) *Learning Lessons from Disaster Recovery: The Case of Mozambique*, Disaster Risk Management Working Paper Series No 12. Washingon DC: World Bank. Team members: Kerry Selvester and Lourdes Fidalgo.

The utilisation of evaluations

Peta Sandison

3.1 Introduction

Although evaluations are generally successful in identifying lessons and building institutional memory, only a minority of evaluations are effective at introducing evident changes or improvements in performance. If this continues, there is a danger that continued poor utilisation will undermine the credibility of evaluation as a tool for accountability and learning in the humanitarian sector.

We do not know even how many evaluations are conducted, let alone how many are used. The source of concern regarding non-use in the sector is mostly anecdotal – from working observations by evaluation managers and users – and proxy, from the apparent lack of impact of evaluation on the sector's performance. After all, if evaluations are intended to improve performance, then 'the recurrence of many of the problems seen in Rwanda and other emergency responses' observed in the synthesis report of the Tsunami Evaluation Coalition (Telford and Cosgrave, 2006, p23) implies that evaluations are not doing their job.

Evaluations of humanitarian action were rare in the 1980s. Twenty years on, the number has soared and evaluation has become one of the most visible features of the learning and accountability agenda. The growth continues, complemented by innovations such as joint and real-time evaluation, participatory evaluation and peer review. We have become increasingly interested in evaluating our performance, but are we using what we find?

The utilisation of research and evaluation has been a topic of lively debate in the development and public sector since the 1970s. ALNAP's humanitarian membership has long been concerned about utilisation, first commissioning a study on the follow-up to evaluations five years ago (van de Putte, 2001). The concern persists.

The results of studies by humanitarian agencies and donors on the use of their evaluations are mixed. Descriptions of limited or absent use are more easily found than examples of good practice. In general the literature describes an inconsistent and, in some cases, a dismal record of evaluation use. A Sida study concludes that 'evaluations are useful to a very limited group of stakeholders. For a majority of stakeholders the evaluation process could just as well have been left undone' (Carlsson et al, 1999, p54).

But others are more upbeat. WFP's study on its follow-up to evaluations carried out over a two-year period concludes that 88 per cent of the recommendations had been implemented or were in the process of implementation. Some two thirds of the recommendations had led to improved performance, although the reviewer recommends caution, noting that 'management units were almost always assessing performance intuitively' (WFP, 2005, p7). A 2002 study on the use of evaluations in the European Commission found that, while the degree of use varied, 'in no case were evaluations considered to be not at all useful' (Williams et al, 2002, p12).

The picture of utilisation that emerges from this and other studies is complex and often unexpected. One of the few certainties is that how and why an evaluation is carried out significantly affects the likelihood of it being used. The studies that constituted the background reading for this chapter provide a significant amount of information about factors that promote the utilisation of evaluations. Much of it resonates with van de Putte's 2001 study, and key references such as Michael Patton's *Utilization-Focused Evaluation*, published in 1997.

Information on use-promoting approaches is clearly available, but is the humanitarian sector using it? And if not, why not? Does the main issue continue to be the quality of evaluations (content and process), suggesting that the evaluation community itself is not learning? Or are there other issues that undermine even the best evaluations? Given the position of evaluation as a primary tool for accountability and learning, is it the right tool for the job? And what kind of job do we expect it to do?

3.1.1 Structure of this chapter

This chapter has five main sections. Following this introduction (Section 3.1), Sections 3.2 and 3.3 draw upon existing studies of utilisation to describe the different types of use made of evaluation, and examine a range of factors found to promote use. Section 3.4 examines four case studies and the findings of a series of interviews and a questionnaire survey. These findings are used to expand on the evidence of the preceding studies, exploring the extent to which factors affecting use identified in the literature are borne out in practice. Section 3.5 considers the implications for the future of learning and accountability mechanisms in the sector.

Box 3.1   Sources of information for this chapter

This chapter draws upon a literature review, four case studies of evaluation utilisation volunteered by CAFOD, MSF(H), OCHA and USAID, semi-structured interviews with 45 evaluators, evaluation managers and evaluation 'users', a review of 30 sets of terms of reference, and an electronic survey sent to ALNAP Observer and Full Members (19 evaluation managers and 27 evaluators responded to the survey).

The researcher was supported by an advisory group composed of representatives from the British Red Cross Society, CARE International, MSF(H), the Netherlands Ministry of Foreign Affairs (Policy and Operations Evaluation Department), OCHA and ODI, plus an independent evaluation consultant and the ALNAP Secretariat.

3.2 What is utilisation?

Utilisation in practice is complex and by no means static. It can mean different things to different users and is strongly related to context. In many cases, utilisation (or 'use' – the two terms are used interchangeably here) is impossible to measure and difficult to attribute to an evaluation.

3.2.1 Types of utilisation

Examples cited in the literature demonstrate a range of expectations of evaluation use. Patton (1997, p 76) describes three primary uses of evaluation findings:

1   judging the merit or worth of a programme (eg accountability to stakeholders; to inform funding decisions)

2   improving a programme (eg ongoing learning and development)

3   generating knowledge.

The first and second uses above are based on a cause-and-effect model. They are both intended to lead to direct changes and decisions. This expectation of use is often referred to as 'instrumental': an evaluation's findings and recommendations should lead to related actions such as tangible changes in policy, funding, systems or operational practice. Many – perhaps most – humanitarian evaluations fall into this category of instrumental use. Evaluations commissioned by donors at the end of a programme or partnership cycle, audits, mid-term reviews, real-time evaluations and so on may have different users and emphases but they share the same expectation of utilisation. They all assess merit, identify strengths and weaknesses and provide recommendations on what to do as a result. Utilising the evaluation means taking its advice.

The expectation of instrumental use is an understandably common one. Instrumental use makes good sense; a set of recommendations provides tangible targets that can, in theory, be measured and monitored. Instrumental use is coherent with the humanitarian sector's current emphasis on results-based management. It is, after all, reasonable to anticipate that the utilisation of an evaluation should be easily recognised and clearly linked to the findings. However, most studies of utilisation conclude that instrumental use is not the norm.

Since the 1970s, the neat, linear connection between evaluation (or research) findings and policy development or programme improvement has been increasingly challenged. 'Evidence suggests that government officials use research less to arrive at solutions than to orient themselves to problems' (Weiss, 1977, p534). Evaluations can influence the user's thinking and enhance understanding, contributing to the development of knowledge (Patton's third category of use, which is also often referred to as 'conceptual' use). Even if an evaluation's recommendations are ultimately rejected, the dialogue, dispute and debate over an evaluation's findings can generate increased clarity and new, more objectively grounded solutions. Such knowledge may or may not lead to action. Any action that does result is not direct and neither does it occur in the short-term. This type of knowledge-generating use is incremental and cumulative. It is typically associated with organisations and policy making, rather than with individuals and specific projects.

Evaluation syntheses might lend themselves best to conceptual use. Over time, the lessons learned and the repetition of key findings drawn from successive evaluations generate new information and subsequent dialogue about patterns of performance at an organisation- or sector-wide level. The utilisation of a synthesis

study is often a result of recognising recurring and therefore potentially systemic issues. The cumulative effect increases the status of the findings as a credible source of organisational knowledge, particularly for senior managers and policy-makers. In instrumental terms, the contributing evaluations may not have been used at all.

Unpredicted and unexpected uses, while not a separate analytical category of use, are certainly worth noting. Several examples were given of evaluations thought long forgotten, later being resurrected or actively used, without the knowledge of the authors or evaluation managers. This reinforces the general finding that the types of use that actually occur are not necessarily those that were planned, nor do they manifest in a predictable form.

Studies of use have also found that individuals and organisations can learn through the process of an evaluation, irrespective of the findings. An extensive review commissioned by the American Evaluation Association (Shulha et al, 1997) examined 74 studies of evaluation use. The review found an increasing prevalence and recognition of such 'process' use, such as participants' increased ownership of evaluation findings, greater confidence in the evaluation process and in applying the results and evidence of personal learning, all of which combine to produce a net subsequent effect on programme practice.

Patton (1997) further notes that users' participation can enhance an individual's commitment to learning and help build a stronger organisational learning culture, contributing to an organisational development process. According to Patton, a user's engagement in an evaluation can also lead to enhancement of shared understanding within teams and the development of monitoring and evaluation as an integral component of programme delivery. An example of this can be seen in a recent joint NGO evaluation in Niger when the evaluation process 'set a precedent for greater collaboration, sharing of resources and information and learning among agencies, and provides them with the opportunity to develop trust, and to regard one another not as competitors but as partners' (Wright and Wilson, 2006, p33).

Another type of use results when an evaluation confirms, rather than informs or creates, a position or knowledge that one or more key stakeholders already hold. This differs from misuse (which is covered below) in that the findings have not been manipulated or distorted to serve the user's purpose but have arrived independently at the same conclusion. A criticism of evaluations is that they often fail to discover anything new. This is particularly pertinent to field workers, frustrated at pages of

familiar information. For stakeholders seeking independent confirmation and the consolidation of information, this is not necessarily a problem. The evaluation is used as an objective source of evidence (to support, for example, the completion and closure of a programme), or as a means of communicating a stakeholder's position. The findings are already known to at least some of the key stakeholders.

A study of the utilisation of 10 MSF(H) evaluations (van de Putte, 2000) notes that the agency had often used evaluations to close an internal debate. Evaluations 'forced a decision'. In all cases, the findings corroborated the existing position of some stakeholders and, inevitably, lost the argument for others. The Sida study (Carlsson et al, 1999) observes that one of the evaluations was clearly intended to communicate the donor's concerns about gender and participation to its partner. The evaluation's subsequent identification of weaknesses in these areas was therefore nothing new to Sida, but it 'legitimised' its position and provided an independent reference for future discussions about programme approach, albeit in an expensive fashion.

Box 3.2 summarises these four principal types of use drawn from the literature, in particular from Carlsson et al (1999), Williams et al (2002) and Patton (1997).

Box 3.2	Types of use

Instrumental use   Direct implementation of the findings and recommendations by decision-makers, leading to related decisions such as future funding, improvements to a programme or changes in policy and procedure. Evaluations that anticipate instrumental use include ex-post, accountability-judgement evaluations or audits and learning-improvement evaluations such as real-time or mid-term evaluations.

Conceptual use   Evaluation results and conclusions trickle down into the organisation in the form of new ideas and concepts debated and developed over time. This type of knowledge-building use is sometimes referred to as 'enlightenment' use (Weiss, 1977). The effect is incremental: single evaluations rarely lead to direct changes in policy and practice but add to, clarify and develop knowledge. Conceptual use is also cumulative – findings may act as a reminder of

CONTINUED

knowledge. Conceptual use is also cumulative – findings may act as a reminder of what was known but previously put aside. Ex-post and evaluation syntheses naturally incline towards conceptual use. However, any type of evaluation or review can contribute relevant findings that generate cumulative knowledge.

Process (learning) use   Participation in the evaluation itself can lead to individual learning and changes in behaviour, such as improved communication within teams and between partners, enhanced understanding and application of M&E in programming. Engagement in the process can also increase users' ownership of evaluation findings and the confidence to use them.

Legitimising use   The evaluation legitimises – confirms, substantiates, corroborates – a decision or understanding that the organisation or individual already holds, providing an independent and objective reference that may be used to communicate or justify subsequent actions.

3.2.2   Misuse of evaluation

Non-use can occur for rational, unintended or practical reasons such as shelving a poor-quality evaluation, bad timing or unexpected events. Misuse, by contrast, is intentional and unrelated to the quality of the evaluation itself. Misuse occurs if an agency commissions an evaluation with no intention of acting upon it, or if an evaluation is carried out as a type of ritual, administrative routine or public-relations exercise.

Accountability is not satisfied simply by transparently identifying performance issues. It also involves a commitment to respond to the findings of an evaluation. Ritual evaluation means that 'an account' has been made (transparency), but no action follows. The number of such evaluations is unknown; if they are common then 'doing' evaluations has become more ingrained, while accountability has not. Such symbolic or ritual use of evaluations can be a cynical manipulation of an agency's public image. It may also be more banal; staff might simply be responding to their institution's administrative directives. The evaluation box is ticked. Either

way, money has been wasted and opportunity costs incurred through the diversion of staff time and energy.

Misuse also occurs when stakeholders deliberately attempt to subvert the independence and integrity of the evaluators and their findings. The Sida study (Carlsson et al, 1999) describes the 'hints' and subtle understanding that can be communicated to evaluators to serve a hidden agenda. While something of an exception in its refreshing honesty, Sida's experience of misuse is unlikely to be exceptional.

Box 3.3	Types of non-use or misuse

Ritual use Evaluations serve a purely symbolic purpose, representing a desirable organisational quality such as accountability. Evaluations are a formality and 'use' equates with the fulfilment of legal or institutional obligations, rather than of the evaluation findings.

Mis-use Suppressing, subverting, misrepresenting or distorting findings; co-opting evaluators to serve a biased agenda for political reasons or personal advantage. Rejecting findings because they do not correspond with the beliefs of key stakeholders or with decisions already taken (such as to cease funding).

Non-use The evaluation is ignored because users find little or no value in the findings (a rational response to a lack of quality or relevance), are not aware of the results (dissemination problem), or the context has changed dramatically (eg evacuation; unexpected closure of a programme).

These different uses and misuses are not mutually exclusive. Components of the same evaluation can lead to a mixture of uses at different times, partly related to the nature of the findings and partly to the users. Different users will select how, and if, they use the findings according to their position, power and interests.

3.2.3 Types of user?

An evaluation useful to headquarters staff in quest of organisational knowledge may not be valued by field staff seeking direct changes in their programme. Individuals within a team or department may also assess the utility of the same evaluation differently. A small number of chief executives were interviewed for this study to gain insight into how such key, high-profile stakeholders use evaluations. While only indicative at best, their descriptions are consistent with the types of use described above.

Box 3.4	Types of use by senior executives

- to support what you want or need to do: independent corroboration or a tool to influence (ie, the findings confirm the prevailing policy direction or individual agendas)

- an independent reference used to resist a directive from above (eg trustees, the board)

- as an outward demonstration of accountability

- to provide guidance when you do not know how to solve a problem

- as an advocacy tool

- as institutional memory.

The evaluation uses cited by the senior executives reflect the needs, interests and imperatives of leadership positions. The literature examined has not explored different types of use by a range of different users, although some logical assumptions could be drawn. For example, ALNAP's research on field-level learning (ALNAP, 2004) noted that field workers are more inclined towards learning from peers, rather than from knowledge contained in reports, guidelines and evaluations. The findings also indicated that field workers prefer learning by doing. If field

workers use evaluations at all, it could be expected that instrumental and process use would dominate, leading to immediate programme changes and learning through participation in the evaluation.

3.2.4   Defining and measuring utilisation

How an evaluation is used appears to defy reliable prediction and seems subject to many factors beyond the control of commissioning agencies or evaluators. The uses are also rich and more diverse than may be expected. If it is common that 'only direct instrumental use of findings and recommendations are regarded as "proper"; use' (Williams et al, 2002, p56), we are failing to recognise the many dimensions of utilisation and therefore doing evaluation a disservice. The picture would look a lot brighter if we accept the recommendation of the European Commission study that '[t]he indirect use of evaluations – including process use, indirect use and cumulative use – should be valued more explicitly' (Williams et al, 2002, p56).

However, such recognition of indirect evaluation use is a mixed blessing. How then do we define utilisation overall and, more importantly, how do we measure it? After all, a poor evaluation could escape censure on the grounds that its utilisation was non-instrumental – diffuse, indirect and therefore hard to discern. If we assess the performance of evaluation in terms of utilisation (as well as its conformity to good evaluation practice), how do we take account of genuine indirect use and learn how to serve it better?

If, for example, use equates with learning, then evaluation should be measuring its performance with respect to learning objectives. But measuring learning is a significant challenge. It is further complicated in evaluation because there is no baseline to compare with, the participants have not necessarily made any formal or informal contract with the evaluators to learn (such as making time for it), and detailed objectives of learning are rarely defined in the terms of reference (ToR) or in any explicit way during the evaluation. The evaluators themselves are often on a steep learning curve, absorbing vast quantities of information about the organisation, programmes, country and individuals and hence are sometimes poorly placed to judge how much of the information generated by an evaluation is new to the users or just new to them.

The difficulties are compounded by variable recall on the part of potential users. Weiss comments that 'the policy maker himself is often unaware of the sources of his ideas. Bits of information seep into his mind, uncatalogued, without citation' (p534). This is not to say that learning is not happening, or that 'learners' wilfully conceal their sources (although professional rivalry and defensiveness may sometimes account for this). Rather, the process of storing and recalling knowledge is complex, and its translation to action is a highly individual and personal process during which individuals transform knowledge and make it their own.

Furthermore, even measuring direct utilisation is problematic. It presupposes that the recommendations are the best or only solution and therefore should have been implemented. Humanitarian and development evaluations share the complexity of the work they are evaluating. Despite the increasing acceptance of sector-wide standards and protocols, evaluation, like programming, is at best a set of informed judgements made in contexts different from those of the events and activities being assessed. There is no gold standard. A dearth of impact data further leaves the door open to different opinions regarding the most effective approach, even in technical domains bound by more widely accepted indicators of good practice. As a result, the recommendations of even the 'best' evaluation can be disputed or rejected on perfectly rational grounds. This further confounds our ability to employ utilisation as a quantitative performance indicator for evaluation.

In summary, an evaluation leads to different types of use depending on the nature of the findings, the interests and needs of the users and the context. Use is neither wholly predictable nor objective. As we shall see in the following sections, there are many identifiable factors that can positively and negatively affect utilisation. Some, though by no means all, of these factors can be controlled or influenced by how an evaluation is designed and implemented. The presence or absence of many of these factors may provide a proxy method of assessing the performance of an evaluation in terms of its potential for utilisation. To do so requires a different definition of utilisation.

It is possible to define utilisation in qualitative terms related to an evaluation's *potential* for use, rather than struggling with definitions that exclude the richness of indirect use and ignore the complex mix of art and science that characterises evaluation use. Utilisation has occurred if the evaluation has led to relevant objective or subjective change as perceived by the users. Drawing on definitions found in

Patton (1997) and Carlsson et al (1999): an evaluation has been utilised if users with the intention and potential to act have given serious, active consideration to its findings, identifying meaningful uses according to their own interests and needs.

The next section explores the factors identified in the literature that promote, enhance and inhibit the use of evaluations.

3.3 Factors affecting utilisation

Why are some evaluations used while others are not? Studies on use have identified a number of influential factors. They can usefully be organised into distinct, though inter-related, groups to provide a framework through which we can analyse and perhaps predict the potential of an evaluation to be used.

The RAPID (Research and Policy in Development) Framework developed by ODI examines the importance of four dimensions that influence the impact of research on policy: the quality of the evaluation information, the political context, the links between the evaluators, policy-makers and other networks and, finally, the influence of the external environment. The RAPID Framework was applied to examine the influence of one of the most famously utilised humanitarian evaluations, the Joint Evaluation of Emergency Assistance to Rwanda (the JEEAR), on the development of the Sphere Project (Box 3.5).

The RAPID Framework provides a useful starting point for grouping the findings of the various studies on evaluation utilisation. For the purpose of this study, these have been grouped into four areas of factors affecting utilisation.

1   **Quality** of the evaluation process and product: the evaluation design, planning, approach, timing, dissemination and the quality and credibility of the evidence.

2   **Organisational** culture and structure: learning and knowledge-management systems, structural proximity of evaluation units to decision-makers, political structures and institutional pressures

3   **Relational** factors: relationships and links between the evaluators and users and their links to influential stakeholders, including shared backgrounds and common ground; relationships between key stakeholders; networks and links between the evaluation community or team and policy-makers

4   **External** influences: the external environment affects utilisation in ways beyond the influence of the primary stakeholders and the evaluation process. It includes indirectly involved stakeholders (not direct users) whose actions can affect the use (or non-use) of an evaluation. These include the public or media, governance structures (eg board, ministers, parliament), executive committee and donors.

These four areas are described in turn below.

Box 3.5	The RAPID Framework and the utilisation of the JEEAR evaluation

'How the Sphere Project Came into Being' (Buchanan-Smith, 2005) notes that the quality of the Joint Evaluation of Emergency Assistance to Rwanda (the JEEAR) was high and the findings actively promoted, targeted and disseminated. There were persuasive and key individuals championing its use, skilfully creating or exploiting alliances and networks. The evaluators themselves were skilled at negotiating the political landscape, identifying key individuals and groups to influence.

The evaluators also had strong links – common interests and close contacts – with the policy-makers. Many of the evaluators had direct humanitarian experience, enhancing their credibility and the legitimacy of the findings with the users. An important role was played by NGO umbrella organisations such as Interaction and ICVA. Specific individuals also had a powerful influence.

Finally, the external environment – the shock of Rwanda, the media spotlight – added to pre-existing concerns about performance in the sector and 'expanded the boundaries of what policy-makers would accept as necessary change'.

3.3.1   Quality factors

Of the four groups, factors related to evaluation quality receive the most attention in the literature on utilisation. This is not simply a question of the quality of the report, at least not in the usual sense: both the product and the process determine quality. How the evaluation is designed, planned, conducted and managed by client and evaluator matters just as much as the report, sometimes more.

Purpose and design

Section 3.2 above described instrumental, conceptual, process and legitimising types of use and the relative value placed on different uses according to the user's interests. According to Patton, the purpose, approach, methodology and presentation of an evaluation should derive from the 'intended use by intended users' (Patton, 1997). One size does not fit all, and 'each purpose privileges different users' (Williams et al, 2002 p35). 'Evaluations are situated in the context of various stakeholders' interests. They are by nature political' (Carlsson et al, 1999, p54). To optimise utilisation, the interests, needs, influence and power of the users should be identified at the outset and used to inform the design of the evaluation.

Patton's insistence on designing around 'specific, identifiable people, not vague, passive audiences' (Patton, 1997, p 382) presents an obvious challenge to evaluations intended for multiple potential users, such as interagency evaluations. He recommends a negotiated prioritisation of users and uses. Evaluations often include mixed objectives and multiple stakeholders, without prioritising or considering what this may mean in terms of approach. If an evaluation is intended to be used to improve an operational programme, then the primary users should be the operational staff. An analysis of their needs and interests would form the basis of the design which may, for example, emphasise beneficiary feedback and certain programme aspects above others. Ownership of the findings by the operational staff responsible for using the evaluation would, in this case, be essential; the written report, less so.

Alternatively an evaluation intended to inform organisational strategy would need to identify the stakeholders with the power and influence to effect changes at this level, and the evaluation should be designed with their needs in mind. The type of information and reduced level of programme detail that would serve their purpose would be of less use to operational staff.

Multiple purposes may displace each other by virtue of overload in the evaluation's scope, and one purpose may unintentionally undermine another. Accountability for one set of stakeholders may displace the intended learning use for others. For example, from the point of view of those whose work is being evaluated, the knowledge that judgements will be made and communicated in writing can create defensiveness. This is in direct opposition to the safe, trusting, non-judgemental atmosphere that modern approaches to adult learning strive hard to achieve.

At the same time, van de Putte's, 2001 study found that respondents differed in their experience of combined accountability and learning uses; some saw no contradiction. What appears to be essential is clarity and agreement about the purpose. The lack of this leads to 'everyone expecting that the evaluation would serve their individual information needs' (Williams et al, 2002, p20); some or all the users will be disappointed. Contradictory assumptions about use are also created by unexplained language. Terms such as learning, accountability and transparency are used loosely and often interchangeably.

Participation

Participation is fundamental to ensure that an evaluation's purpose and design are relevant to the users. The 'active participation of stakeholders at all stages in the evaluation cycle' (Williams et al, 2002, p53) promotes use. Participation is not passive listening or the act of being consulted. It demands meaningful and sustained involvement from those given the potential to make decisions about the evaluation process. 'High-quality participation is the goal, not high-quantity' (Patton, 1997, p383). Exposure to the process and learning also means that participants are 'in the best position to internalise these lessons' (OECD, 2004, p27) and more likely to value and trust the findings (Ridde and Sahibullah, 2005). Forms of participation depend on the context; they can include steering committees, regular group or individual discussions between the evaluators and users, workshops and advisory groups.

Participation throughout an evaluation also reduces the potential bombshell of a critical assessment or the 'shock' at recommendations for significant change. Individuals and organisations are more disposed to change if they are familiar with the information and mentally prepared. Studies often mention the low involvement of senior managers during an evaluation. Their involvement, if at all, occurs often only at the report stage. This may be significant for the low utility ascribed to evaluation by senior managers in several of the studies assessed. Sudden exposure

to proposed changes that are complex and demanding increases the risk of wholesale rejection or its equivalent – indefinite postponement.

Planning

Good planning also supports utilisation. Allowing adequate lead-time facilitates appropriate recruitment, wider consultation in the design stage and a good fit between the evaluation and key decision dates such as funding cycles and annual programme planning. A study of MSF(H)'s use of evaluations (van de Putte, 2001) notes that in several cases evaluations were not used because they took place too late: the decisions they should have influenced had already been made. Timeliness plays a particularly important part in humanitarian evaluation, particularly in field-based evaluations of short-term emergency programmes and RTEs when the window of opportunity is particularly brief.

Quality of the evidence

A great deal of emphasis has been placed on the content and form of evaluation reports. ALNAP annually assesses the quality of evaluation reports with respect to a set of criteria developed in its Quality Proforma. While studies of use have demonstrated that the (written) product is not sufficient in itself to guarantee utilisation, it is nonetheless an important factor. If the quality is poor, data dubious and recommendations irrelevant, it is perfectly rational not to utilise an evaluation. Although methodological rigour is important to ensure that collected evidence is credible, this is less of an issue for users than is the accessibility and relevance of the report. Time and again, excessive length and inaccessible language, particularly evaluation jargon, are cited as reasons for non-use.

Feasible, specific, targeted, constructive and relevant recommendations promote use. This is partly related to the user-focus and participation discussed above. Relevant, viable recommendations are more likely to be conceived if the users have been involved throughout the evaluation, communicating institutional and individual needs. Low relevance may explain why several studies noted that senior managers were the least likely to use evaluations. Although this is partly related to organisational and political issues, discussed below, it may also be similar to the experience of managers in the Finnish study who find that 'evaluations are often too backward-looking, they tend to say things with the benefit of hindsight. This kind of evaluation is seen as contributing very little to the management's work which essentially consists of finding future orientation' (p 11).

This resonates with the profile of senior managers described in the personality profiles such as Myers Brigg analysis as people who are future-focused and problem-solving by nature. Evaluations that do not provide some anchors for positive action because there is a critical rather than constructive problem-solving emphasis are less likely to be used

Some writers think that it is not the evaluator's job to make recommendations. An evaluator should generate the findings but it is the users' job to translate them into recommendations or actions. Anyway 'evaluators seldom have the expertise to make recommendations' (Patton, 1997, p327). Patton believes that the solution is the generation of recommendations as a collaborative process

Mechanisms for follow-up

How an evaluation is disseminated and followed up is another important utilisation factor. This entails far more than the distribution of the report. 'Many of our cases have shown that the evaluation report is over-emphasised, at the expense of other ways of disseminating findings' (Carlsson et al, 1999, p45). Dedicated follow-up individuals, the clear allocation of responsibility and specific mechanisms for action increase the likelihood of evaluation use, particularly if follow-up was planned from the beginning of the evaluation.

Examples of follow-up mechanisms include management action plans, follow-up working groups, evaluation communities and networks, the European Commission's *fiche contradictoire* (which invites a systematic response from potential users) and the creative use of the internet to disseminate findings and invite consultation. Participants at an OECD workshop in Tokyo (OECD, 2001) shared innovations such as CIDA's development of pop-up reminders for staff that appear at key planning and approval moments, linked to a learning database. The World Bank is investigating the utility of types of search engines based on linguistic analysis, capable of 'zeroing in on relevant material' (OECD, 2004, p 30). Other studies found that the incorporation of evaluation findings into training programmes enhanced use, although this is relatively rare and under-resourced.

An important consideration is matching the form of dissemination to the users. There is some evidence to suggest that adapting the mode of dissemination to suit the user is a stronger predictor of utilisation than designing the evaluation around their needs in the first place (Landry et al, 2001). 'Careful targeting allows intelligent tailoring of feedback approaches' (OECD, 2001, p 22). Danida for example has

identified three audience groups and uses different formats to suit different users, including a shortened 'popular version' of a report, newsletters, lectures and seminars. Users in some studies also requested reports that were customised for content, not just length, identifying the particular interests of, for example, senior managers and extracting relevant strategic issues. This, however, appears more likely to occur through verbal debriefs, and several of the OECD agencies emphasised the continuing value of face-to-face communication.

The evaluators

Although some of the studies (eg van de Putte, 2004) note that finding experienced evaluators is often a problem, surprisingly little is said about the competence of the evaluators. Their role with respect to utilisation is also notable by its absence. It is rare to find reference to the use of evaluators as a follow-up resource or trainers, or to any involvement of evaluators beyond the final report and workshop.

The evaluators themselves may not regard utilisation as their responsibility. The work predisposes consultants to an itinerant and potentially piecemeal approach. Patton departs from the traditional notion of evaluators as a type of researcher detached from the operational – and political – realities of utilisation. He argues that the independence and integrity of an evaluation is preserved through adherence to professional standards and guidelines, not through disengagement from the users' needs and context. He considers that evaluators have the responsibility to facilitate use, secure the attention of key users and understand how to make evaluation information compete successfully with other organisational imperatives. Further, evaluators should 'train users in evaluation processes and the uses of information' (p383).

To do this would require adequate time allocation for such quality interaction. Little is said in the literature about this, and the extent of such involvement in practice is difficult to ascertain from reports. For evaluators, assessing the potential for a utilisation approach before starting is easier said than done. Essential preconditions, such as the client's degree of motivation and availability to engage in the process, are hard to discern at the contractual stage. Proxy indicators, such as difficulty in obtaining access to key staff members for interviews, may quickly reveal the real engagement of stakeholders in the process, but by then it is too late; 'it's better to find out before preparing the meal that those invited to the banquet are not really hungry' (Patton, 1997, p85). If evaluators seek to ensure that their work is to be

used, far more pre-evaluation engagement (including participation in the development of the ToR) would be necessary.

Clients cannot judge the ability of an evaluator or team to facilitate utilisation, and the general absence of accreditation means that competencies are already varied and difficult to predict. The recognition of conceptual and process use implies a range of evaluation skills that exceed fluency with traditional evaluation methodology. 'A need exists to rethink the professional profile of evaluators that are to engage in enhancing learning. Next to research skills, such professionals should excel in communication, process facilitation, conflict resolution and negotiation skills. Analytical skills and tools will have to include situation and stakeholder analysis, as well as group dynamics' (Carlsson and Engel, 2002, p9).

The array of quality factors affecting use supports the importance attributed by the evaluation sector to good practice. At the same time, evaluation does not take place in a vacuum. Its use is also a function of the links between the evaluators and managers and the users.

Box 3.6	Summary of quality factors promoting utilisation

Design – purpose and approach   The purpose is clarified and agreed among the key stakeholders at the design stage; the potential for multiple, potentially contradictory purposes is avoided through discussion and prioritisation. The approach is designed in accordance with the purpose and the users' needs and interests (eg a field- learning or programme-improvement purpose should ensure maximum involvement of the field staff in the design and a reduced emphasis on other stakeholders' needs such as for reports).

Participation and ownership   There is meaningful, quality participation of all key stakeholders throughout the evaluation. Users directly influence the purpose and design, increasing its relevance and ownership, and are appropriately involved throughout the process (through clear mechanisms such as steering groups, discussion, clarification, regular feedback throughout the process).

Planning   There is sufficient lead-time and adequate allowance for quality staff investment in the evaluation. The evaluation coincides with key decision-making

CONTINUED

cycles or events such as a new programme cycle, new budget year/allocation or parliamentary debates. Unplanned timing can also be a factor – such as if the evaluation coincides with an unforeseen opportunity, event or crisis that increases its relevance. Lastly, timeliness is important. The utilisation window of opportunity is particularly short for operational humanitarian programmes and easily missed by relatively minor delays.

The evidence   The evidence is credible (well researched, objective, expert). The report is easy to read (concise, with accessible language and no jargon). Recommendations are specific, prioritised, constructive, relevant, feasible and identify who is responsible for action and when.

Follow-up mechanisms   Specific follow-up plans are established at the outset: clear allocation of individual responsibilities; a 'champion' or key person is committed to action; formal management response mechanisms and action plans; pro-active dissemination through promotion of the findings in user-relevant ways (eg through management meetings, presentations, seminars, team discussions, fast-track debriefs) as well as wide distribution through, for example, the internet. Use may be enhanced through customised dissemination, with content as well as form extracted or adapted for specific users (usually senior staff members).

Evaluator credibility   The evaluator is credible (in terms of competence and reputation). The ability of the evaluator to be balanced and constructive is also important; wholesale negativity tends to lead to wholesale rejection. The evaluators and evaluation managers understand the political nature of evaluation, facilitate utilisation and manage stakeholders accordingly. A broadening of skills to include facilitation, stakeholder analysis, the management of conflict and group dynamics are important for user-orientated evaluation.

3.3.2   Relational factors

Efforts to ensure the independence of evaluators and evaluation units are essential to protect the credibility of the findings but can inadvertently undermine use. 'Some

agencies said that it is fundamental to the role of evaluation units, and is in fact their key attribute. Others argued that it can be a hindrance if independence becomes a barrier to partnership and dialogue' (OECD, 2001, p39). Independence can lead to a perception that evaluation is too far removed from operational and organisational realities. If evaluation is regarded as a somewhat academic exercise carried out by specialists, it can be viewed as opaque and resistant to users' needs. 'The motivation with regard to evaluation has been extinguished because everything in that respect comes in a ready-made package from the evaluation unit with the feeling that we have no influence whatsoever over it' (MFA Finland, 1997, p17).

Landry's study of research utilisation in Canadian public administration found that one of the best predictors of use is the intensity of the linkages between the researchers and policy-makers. The ability of individual evaluators to establish relationships based on mutual respect and trust also appears to be important. This is not to say that relations must necessarily be particularly friendly or informal. Cordial and professional relations between evaluators and clients are realistic expectations and appropriate for the job. At the same time, as noted by the Sphere case study (Buchanan-Smith, 2005), common ground between the evaluators and the policy-makers is important; the JEEAR evaluators came from the same 'community', establishing 'chains of legitimacy' (p102) and credibility.

Attitudes on both sides that generate a sense of two cultures or communities are unhelpful. Relational issues are sensitive topics and of course do not appear in the products of evaluations to permit review. This may be an important step to consider in the management of evaluations, and to help build trust in the process.

3.3.3 Organisational factors

Evaluation as a tool for organisational change

Individual learning contributes to organisational learning. 'But since organisations are not like individuals – and should not be anthropomorphised – they "learn" through processes that reflect organisational interests, structures, functions and decision-making context' (Suhrke 1999, p70). The complexity of organisational change means that it is hardly surprising that evaluation's contribution to the process is at best indirect. Recommendations aimed at policy or structural changes

> ### Box 3.7   Summary of relational factors promoting utilisation
>
> **Personal and interpersonal**   The evaluator is able to establish constructive relationships with key users. Trust is established. The evaluator's interpersonal skills and commitment to quality can be important. Commonality of background and skills between evaluator and users enhances credibility. Overall, the perceived credibility of the evaluator is important.
>
> **Role and influence of an evaluation unit**   The evaluation unit manages to maintain its independence from decision-makers while ensuring close integration and relationships. The unit is able to play a mediating role between stakeholders. A key feature is to establish common ground and shared relevance between evaluation, policy and programmes through communication and relationship building.
>
> **Networks, communities of practice**   Key stakeholders (users or evaluators and evaluation units) have links with broader networks, influential fora and individuals, which enhances the credibility of the evaluation, extends its reach and ability to influence wider policy

in an organisation imply consensus and action from a sometimes-fantastic number of stakeholders. Several of the studies noted that findings that challenged strongly held beliefs and behaviour embedded in an organisation's culture were less likely to be implemented than technical fixes and discrete units of change. Utilisation is also higher when an evaluation dovetails with other initiatives for change. Even the highly influential JEEAR was most successful where it added momentum to existing initiatives.

A paper on the future of humanitarianism argues that organisations seek, above all, consistency and predictability, and shy away from complexity:

> The more a matter is seen to offer an immediate and direct solution to problems at hand, the more it catches the attention of those who determine policies and make decisions. Conversely, the more ambiguous and less

immediate the information, the greater the divide between the strategist, the planner and the decision-maker. (Kent, 2004, p8)

This paper also notes that, faced with lessons learned that apply across the organisation, humanitarian agencies have a tendency to divide them into smaller, manageable chunks that can be packaged out to individuals and departments. This behaviour may partly explain the attraction of the matrix approach to evaluation follow-up. Though efficient and good for individual accountability, the division of labour swiftly removes the process of intra-organisational cooperation and dynamic debate that is needed for organisational change. The same is true of sector-wide learning.

Obstacles to organisational learning

Van Brabant's study (ALNAP, 1997) describes a number of other obstacles to organisational learning:

centralisation and hierarchy, internal power structures and poor information management… having to live up to the false image that development and relief are quick and easy, the temptation to hype up one's performance in the face of growing competition between agencies, the financial instability of certainly (*sic*) humanitarian aid agencies, caused by a trend to go for cheap growth driven by short term funding and unrealistically low overheads, and the high degree of job insecurity of many staff. (p3)

Cautious leadership and 'the governance structure of a charitable board of trustees tends to exert a centralising and risk-averse influence. Board members (who in many countries cannot be paid for their work) may lack time or experience of development processes or of NGO management' (Mango, 2004, p4). Other impediments described in the literature include an organisational culture that does not value learning, staff members who do not understand evaluation, bureaucratic imperatives such as the pressure to spend regardless of quality, and the lack of real incentives to change.

The unequal nature of the aid relationship is also a significant barrier. Why and by whom an evaluation is commissioned affects ownership and hence utilisation. A paper prepared for an OECD/DAC workshop on evaluation feedback found that if an

evaluation was perceived, rightly or wrongly, as donor-driven, the impact was 'sharply reduced' (MFA Netherlands 2000, p5). The same can be said of evaluations viewed in the field as serving only headquarters needs, not the needs of the programme.

Challenges in linking policy with practice

High staff turnover and poor knowledge management are well-known impediments to embedding organisational learning in routine working practices. Furthermore 'evidence shows that NGOs are not tightly managed entities. Senior field staff work with considerable latitude, using their own judgement. Organisational policies are not always followed in practice' (Mango, 2004, p4).

Performance issues can also inhibit use. Just as utilisation is enhanced by motivated individuals willing to 'champion' the findings and promote use, it is also constrained by individuals who block or fail to act. There is some evidence to suggest that evaluations are sometimes used to compensate for the inability of management to persuade staff to adopt new policies and practice, with varying success.

Organisational imperatives

Evaluation reports can present a risk to an organisation's reputation. 'A lot of time and energy has been wasted in dealing with the public out-cries caused by the negative handling of an evaluation by the media', displacing the utility of the evaluation (MFA Finland, 1997, p11). The perceived risk may lead staff members to suppress and reject findings in the interests of protecting their survival. 'The need to maintain a high profile and operate within an increasingly competitive environment encourages a preoccupation with survival and self-justification across the sector, where organizations may lose sight of their higher ideals, focussing instead on activities that ensure their own continuation and growth' (Wigley, 2005, p23).

A study on obstacles to organisational learning notes a number of such 'paradoxes' in aid agencies. In Sida for example:

> Managers were supposed to spend the funds allocated to their programmes and to their countries, but too often it was impossible to spend the money fast if it was supposed to achieve the programme goals,

> whether they were reaching the poor or building sustainable institutions. They were forced to spend money on badly prepared or ill-conceived projects in order to avoid being criticised and seeing their budgets cut in coming years. (Edgren, 1999, p54)

There is also a limit to the absorptive capacity of the humanitarian system. It is often described as under considerable strain. Workload and the increasing number of upwards-accountability demands further reduce the space for learning. Time is a significant issue in the capacity of personnel to absorb learning and address change that requires sustained investment. Referring to understaffing and 'general working conditions approaching the unreasonable' (p18), the Finnish MFA study notes that as a result staff members concentrate on immediate demands; 'evaluation products never have the top priority because they do not require immediate action on the part of the staff member' (p18). The force and dominance of internal organisational imperatives imply that it is easier to discount performance lessons from evaluations than to ignore a threat to personal and organisational reputation.

It is clear that the majority of the characteristics and impediments to organisational change are beyond the control of evaluation. Facilitators and predictors of utilisation in general relate to the presence of positive structural and cultural characteristics that predispose organisational learning. In larger organisations, the existence of a well-resourced evaluation unit is, for example, an important determinant of use. The European Commission study concludes that Directorates-General with more institutionalised evaluation functions (a dedicated unit with significant evaluation experience, skills and established procedures) are more likely to use evaluations. Close structural links between evaluation managers and decision-makers are also significant.

3.3.4 External influences

External pressure may be a significant utilisation factor according to some studies. 'Many authors however feel that most change results from external pressures' (ALNAP, 1997, p3). The question is whether it is (published) evaluations that provoke external pressure. The ODI Sphere case study concludes that the humanitarian system is 'most responsive to change under pressure when the push factors are high' (Buchanan-Smith, 2005, p98), referring to the media and donor pressure

| Box 3.8 | Summary of organisational factors promoting utilisation |

Culture   Senior managers promote a culture of learning (openness to scrutiny and change, embedded learning mechanisms, transparency); staff members value evaluation and have some understanding of the process. Attention to performance is integral to working practice, managers actively support staff to learn and the organisation's leaders promote and reward learning.

Structure   An evaluation unit or individual dedicated to accountability and learning is in place. The evaluation unit is structurally closely linked to senior decision-makers, adequately resourced and competent.

There are clear decision-making structures, mechanisms and lines of authority in place. Vertical and horizontal links between managers, operational staff and policy-makers enable dissemination and sharing of learning. There are permanent and opportunist mechanisms for facilitating organisation-wide involvement and learning.

Knowledge management   The organisation has functional knowledge-management mechanisms, including systematic dissemination mechanisms, informal and formal knowledge-sharing networks and systems.

following the Rwanda crisis. Donors exerted pressure on UNHCR following the Kosovo evaluation, which, according to UNHCR's RTE Afghanistan, directly influenced its programme there (UNHCR, 2004, p14).

Yet there is little evidence in the studies of utilisation to suggest that evaluation is typically used by external stakeholders as a tool to effect change. As there is no regulatory body to monitor performance in the sector, evaluation remains an essentially internal reference without formal power. The recent Humanitarian Impact Conference, 2006 concluded that 'despite the accountability feature of impact evaluations, no consequences exist for organizations or individuals whose programs demonstrate poor performance' (Fritz Institute, 2006, p2). Despite the unusual influence of the JEEAR, its impact 'was also partial. The more challenging and radical recommendations around accreditation and regulation of NGOs were ducked and have been consistently evaded' (Buchanan-Smith, 2005, p102).

The issue of greater regulation and certification has been raised again by the TEC synthesis report, ten years later. The TEC synthesis also notes the absence of real external drivers for change. 'This lack of external pressure for change is one of the critical reasons why performance has improved so little in the last 10 years (including the performance of donors)' (Telford and Cosgrave, 2006, p109).

Protecting reputations and funding

This is not to say that the reaction of external stakeholders to published evaluations does not provoke a response from humanitarian agencies. The *fear* of repercussion from publicised criticism is certainly very real. However, agencies seem most likely to be galvanised to protect their reputations. The rush is not towards the recommendations. To complicate matters further, it does not appear to be performance that influences the funding decisions of either donors or the public. Other influences are at play.

A 2005 report by the UK's National Audit Office examines DFID's engagement with six multilateral agencies: 'Internationally, the effectiveness agenda is rising in prominence but is not yet at the forefront of funding decisions generally. Other donors surveyed felt that effectiveness was just one of a number of issues to consider when deciding who to fund' (NAO, 2005, p23). The ability to shop around is limited, in part, to particular global mandates and the availability of agencies on the ground. Moreover, donors are dependent on their partners and share a similar interest in painting a positive picture of performance. The increased presence of donors in the field has also reduced the power of evaluations as a source of information.

Many NGOs rely heavily on public donations. A loss of credibility through reports of poor performance would surely affect such voluntary funding. However, a recent survey of UK charities found that the public gave to charities based on trust and a belief that the money is spent 'wisely and effectively', despite a 'distinct lack of public scrutiny, and scant knowledge, over how charities are managed' (Charity Commission, 2005, p23). The most persistent public concern regarding charities continues to be the amount they spend on overheads, not their performance (Slim, 2002).

There is little in the studies to advise evaluation managers on how to manage the public's (unrealistic) expectations and the risks associated with transparency. Efforts to educate the public and media are an obvious route and one that some organisations such as the British Red Cross are pursuing. More honesty about the

complex reality of humanitarian work is a necessary but potentially risky step as agencies currently design their funding messages around the public's expectation of a simple and effective response administered by worthy outsiders. Addressing many of the performance weaknesses identified by evaluations (such as greater beneficiary participation) could alter the product–process ratio unfavourably and incur exactly the kind of increased overhead costs so mistrusted by the public.

There are moves to counter the impasse. According to the OECD workshop report, Danida's analysis is that damaging headlines are the result of attempts to conceal and the inevitable lure of an exclusive. In response Danida has increased the availability of evaluation reports and routinely issues press statements backed up by press conferences to permit intelligent debate. Japan has even included journalists in evaluation teams (OECD, 2001).

Pressure from the beneficiaries?

External pressure may come from the public or government of the recipient countries, seeking better accountability for the impact of money spent in the name of their population. 'Tolerance for the evaluation gap is waning. Developing country governments are demanding better information about the efficacy of social spending' (CGD, 2006, p3). This may yet become an influence on humanitarian funding, though the obstacles are higher and the likelihood low in countries with failed states and populations without voice.

The 'customers' should constitute an external pressure. Although beneficiaries are more likely now to be consulted during evaluations, it is not common practice to communicate the findings afterwards.

> Humanitarian agencies are perceived to be good at accounting to official donors, fairly good at accounting to private donors and host governments, and very weak in accounting to beneficiaries. This points to the very heart of the challenge confronted by those who wish to promote humanitarian accountability. The people whose welfare is meant to be the object of the exercise have the least say (indeed, often none at all) in designing policy or shaping operational practices. (HAPI, 2006, p3)

A study examining the motivation of governments to take development evaluation seriously (Gordillo and Anderrson, 2004) notes that: 'producing good quality

evaluations studies alone are not sufficient to motivate political action' (p13) 'The role of beneficiary populations in the evaluation design and implementation seems to be particularly important in creating incentives that support action' (p13) and 'downward accountability was crucial in motivating politicians to take the results of the studies seriously' (p13).

Of course there are numerous reasons why it may not be practical or even feasible to provide beneficiaries in affected communities with the results of evaluations. In some cases the political context means that findings may be too sensitive to share. At a community level, if the beneficiaries have not been active participants in the programme's design, then it is probably too late to engage at the evaluation stage. Mechanisms for feeding back to dispersed and vast beneficiary populations would have to be carefully planned and targeted; they may be impractical (due to problems of access) and could take as much time as the conduct of the evaluation itself. However, the marked absence of any intention to provide downwards feedback deserves further attention. There are examples of good practice such as Action Aid's new Accountability Learning and Planning System (Alps), which emphasises downwards accountability and participatory review. Such initiatives could be encouraged and provide a powerful moral incentive to act on evaluation results.

3.4 Utilisation in practice

This study sought to explore utilisation further through new case studies, a greater emphasis on the experience of the evaluators and particular users such as senior managers and through analysis of some sector-specific datasets such as an analysis of evaluation ToRs and a review of previous ALNAP work in this area. The types of use and factors affecting use described above provided a framework within which to analyse and test the findings. Four case studies were volunteered by ALNAP members for this chapter, specifically from CAFOD, MSF(H) and OCHA, plus a shared USAID–DFID evaluation (Box 3.9).

3.4.1 Types of use

MSF(H)'s use of an evaluation of its nutritional intervention in Bahr el Gazal (Valid International, 1999) is an example of the complexity associated with conceptual use. The Bahr el Gazal evaluation did not, by and large, lead to direct policy or practice changes, at least not at the time. Rather, it was initially rejected by many in the

Box 3.9	The case studies used for this chapter

The **MSF Holland** case study focused on the impact of three evaluations on the development of MSF's nutrition policy and approach: *MSF in Catastrophe* (Valid International, 1999), *MSF Holland in Afghanistan Mission Evaluation: May 2000 – May 2002* (O'Reilly and Shoham, 2002) and *MSF Holland's Medical and Nutritional Strategy in Darfur – Sudan* (Corbett and Lloyd, 2005). Six MSF personnel and the three evaluators of the 1999 and 2005 studies were interviewed for the case study.

In 2005, **CAFOD** commissioned *Rudder and Sails* (O'Donoghue, 2005) to 'Review the strategic priorities of CAFOD's Humanitarian function and provide clear feedback, analysis and recommendations on how CAFOD determines if, when and how to respond to a humanitarian crisis'. Six senior managers and the evaluator were interviewed for this study.

The **USAID Office of US Foreign Disaster Assistance (OFDA) and the UK Department for International Development (DFID) joint evaluation of the Humanitarian Information Centres (HICs)** was commissioned in 2004 to 'ascertain the effectiveness of HICs in servicing the humanitarian community' (Sida and Szpak, 2004). Five people were interviewed for this case study, from OFDA, DFID and OCHA, and including one of the two evaluators.

The Inter-agency Real-Time Evaluation of the Humanitarian Response to the Darfur Crisis 2005 was commissioned by the Emergency Relief Coordinator and Under-Secretary-General for Humanitarian Affairs (ERC/USG), Office for the Coordination of Humanitarian Affairs (OCHA) (OCHA, 2006). Eight interviews were carried out for this case study, with five OCHA personnel, one donor and the two external evaluators.

organisation. However, it generated new data, raised questions and proposed innovations that challenged MSF(H)'s thinking. It led to intensive debate (which continues) and to the development of alternative approaches in nutrition, linked to the ideas presented in the evaluation but adapted to MSF(H)'s culture and mandate.

The evaluation also influenced the evaluator. The inherent limitations of therapeutic feeding centres, particularly in contexts of overwhelming need such as Bahr el Gazal in 1999, led Valid International to develop an alternative approach (Community-Based Therapeutic Care, CTC) which has subsequently directly or indirectly influenced the nutritional approach of most major agencies in the sector.

The initially limited impact of the 1999 evaluation was due in part to the divergence of some strategic recommendations from MSF's organisational ethos and mandate. Several of the recommendations concerned community participation and an improved understanding of local cultures, a significant departure from MSF's medical approach to nutrition and its focus on centralised, expatriate-managed emergency response. The evaluation also predated the full development of a community-based approach; the ideas were in their infancy and difficult to operationalise. Particularly for senior management, the findings challenged (and could damage) MSF's reputation as a highly effective lead in the nutrition-response sector. The evaluation itself seeded a shift in thinking and contributed to a slow process of review and further development. Its 'use' was conceptual, rather than direct and instrumental.

CAFOD's review of its humanitarian function (O'Donoghue, 2005) stimulated new thinking and learning. The review took place within a broader process of organisational change, and has led to significant policy changes. In this sense, the use was instrumental, with action clearly linked to the evaluation findings. However, the recommendations themselves were the product of a highly interactive and iterative process which stimulated, 'downloaded and captured the ideas and thinking of individuals'. In other words, the recommendations did not spark the changes, they captured the dialogue facilitated by the evaluation and evolved from a process of change already in motion.

Through the process of the CAFOD review, staff became aware that they had not really 'understood' the new organisational policy. They 'knew' about it, and had signed up to it, but they had not understood what it meant in terms of operational practice, the capacity, roles and responsibilities of their departments. The review

clarified the implications of the new policy and of a changing context for CAFOD. This process learning also reinforced and developed learning derived from other sources. The cumulative effect was viewed as extremely important as it increased the credibility and force of the review's findings and contributed to longer-term conceptual learning in the organisation.

In 2004, USAID/OFDA and DFID jointly commissioned an evaluation to 'ascertain the effectiveness of HICs in servicing the humanitarian community'. Although there was some instrumental use, USAID, DFID and the Humanitarian Information Centre (HIC) manager commented that there was little in the evaluation that was new to them. The donor's involvement and the ongoing availability of mission reports and reviews meant that the evaluation contained few surprises. However, although DFID would have appreciated more radical recommendations, neither donor viewed the predictability of the findings as a weakness. It led to some instrumental use, but the most important use of the evaluation was by the donors because it substantiated and consolidated existing information (making it easier to share). Importantly, it confirmed and independently validated their perception of the HICs. It also provided a kind of benchmark and reference document about the HICs.

Similarly, MSF(H)'s 2005 Darfur evaluation provided confirmation of what was known or suspected; it legitimised a position. It provided an independent reference for ongoing discussions about the effectiveness of MSF(H)'s performance in nutrition in Darfur and elsewhere.

The case studies also demonstrate non-use. Although USAID and DFID both responded to some of the recommendations, there was little direct impact on OCHA. Some 18 months after the HIC evaluation, the new manager initiated a joint USAID, DFID, ECHO and OCHA review of the new Pakistan HIC, which noted that '[l]essons identified in both the 2004 USAID/DFID evaluation of HICs and the 2005 OCHA/ECHO monitoring mission to Sumatra and Sri Lanka have been only minimally implemented' (p3).

The case studies demonstrate the direct and indirect effects of some evaluations on organisational and individual learning. While some instrumental use is evident, the case studies reinforce the findings derived from the literature overall; use of evaluations is varied, sometimes unpredictable and with particular actions or changes difficult to attribute to a particular evaluation. Conceptual and process use is common, as is legitimising use. Instrumental use is the least likely form of

utilisation, and one evaluation can lead to a mixture of uses, as well as partial or selective use. Commissioning evaluations for purely ritual reasons, such as to meet administrative or legal directives or to symbolise organisational accountability without any real commitment to using the results, may be quite widespread.

3.4.2 Quality factors affecting use

To what extent do quality, relational and organisational factors explain the use and non-use of evaluations in the four case studies? What is the working experience of evaluators and evaluation managers, and what do the ToRs in circulation tell us about the way in which recent evaluations have considered use-promoting factors? The following three subsections (3.4.2 to 3.4.4) address these questions.

Design: purpose and approach

A review for this chapter of 30 evaluation ToRs, randomly selected from those distributed through the ALNAP network, shows that a combination of different purposes is common. Of the 30 ToRs reviewed, 22 contained two purposes, most often a combination of accountability (judging effectiveness) and organisational learning. Over a third contained a learning (or programme-improvement) purpose, but rarely in isolation. There was little evidence of variations in the evaluation design according to these different objectives.

Several of those interviewed for this study thought that accountability and learning purposes should be separate. If accountability is sought, then an audit approach would be more appropriate. The intention would be to measure performance against standards and indicators (such as Sphere, organisational mandates or international agreements and policy instruments). The audit would be linked to mandatory cooperation from the 'subjects' of the exercise, rather than through the semi-negotiated access common to evaluation. Learning objectives would become a separate, internal process designed around specific field needs. This would imply outcomes that do not rely on – or sometimes even produce – a written report. Interestingly, several evaluators mentioned the difference between what is said (and agreed) verbally with staff and the shift in perception (and defensiveness) when the same thing is put in writing.

At the very least, a mixed agenda should be explicit. Some of the ToRs reviewed for this study demonstrated a degree of coyness about straightforward accountability objectives; the purpose is often referred to as 'learning', without reference to the principal motivation of the evaluation, which is to judge and report back. This may not be intentionally disingenuous, but the lack of clarity (and courage) about the purpose can generate suspicion towards the evaluation and confusion for the users.

The sector is increasingly experimenting with new approaches to evaluation suited to particular uses. After-action review emphasises internal learning. Joint evaluations can address the specific issues of coordination and collaboration that single-agency reviews do not. Real-time evaluation (RTE) can focus on immediate feedback and learning-in-action that ex-post evaluation cannot.

It is too early to assess whether innovations such as joint and real-time evaluation are being used for their intended purpose or not. The experience of some joint evaluations described in Box 3.10 is mixed. Some of the advantages of joint evaluations have been realised, such as presenting an unbiased common picture of the whole response. Equally, the disadvantages which would be predicted by Patton's user-based approach are evident. Findings are too general, ownership is lower and the evaluation lacks specific stakeholders responsible for subsequent action. Many of these joint evaluations combine a review of interagency process with the sum of fully fledged individual evaluations of each agency. This may lead to an evaluation that is overstretched and misses all its targets, displacing its natural focus on how the system is working as a whole.

The case studies similarly demonstrated intermingled objectives. For example the Darfur real-time, interagency evaluation combined the objectives of immediate improvements to field performance and organisational learning for future humanitarian action. This combination of field and organisational learning objectives meant that the evaluation design had to fulfil immediate learning needs (eg a focus on the field, rapid feedback, an iterative approach) while also providing the type of information needed by senior policy-makers and managers (eg comprehensive reports). Additionally, the public availability of the report and the fact that it was viewed as driven by donor criticism of the response led to a perception in the field that upwards accountability was the main real purpose of the evaluation.

Box 3.10   Joint evaluations

The concept of joint evaluation has been promoted by OECD/DAC since 1991 but the experience has been patchy. A review in 1998 noted a number of limitations, including 'too general and diplomatic conclusions as they have to combine different interests'. It found that there was little enthusiasm for joint evaluations but, encouraged by the GHD initiative, donors have returned to the issue recently. There has been a DAC Evaluation Network Working Paper, a conference in Nairobi and publication of OECD/DAC 'Guidance for Managing Joint Evaluations', updating an earlier paper from 2000.

A review of the Niger evaluation by the participating agencies (CARE, Catholic Relief Services, Save the Children and World Vision) (Wilding et al 2005) concluded that the advantages of this joint evaluation were:

- presenting the 'big picture'
- lack of bias
- provision of a basis for setting up Niger's NGO coordination forum, and provision of direction for recovery activities
- shared costs
- a common picture for donors.

Among the disadvantages in this specific case were:

- lack of a focal point to take forward the evaluation recommendations
- agencies did not participate equally
- different approaches of the agencies, causing conflict within the evaluation team.

Overall, the NGO experience has been mixed. Another joint evaluation supported by the ECB, in Guatemala, was considered less successful by the commissioning agencies. The joint Tsunami evaluations by CARE, Oxfam and World Vision were also considered to be difficult experiences though they did assist each agency to improve specific programme work. The general lesson so far seems to be that joint evaluation is a complicated exercise requiring a level of engagement that stakeholders on the ground are not usually able or willing to provide.

The specific objectives in the Darfur ToR are wide-ranging and include most of the evaluation criteria that may be expected in ex-post evaluations. The overload evident in the ToR was in large part a hazard of a multiple-agency evaluation. The evaluation team also tried, appropriately, to respond to emerging needs in the field. As a result, the evaluation addressed issues as diverse as water quality and the role of the Security Council. Field staff in Darfur viewed the visits as 'dipping in and out' without being able to develop and deepen the analysis. There was little evidence in the case studies of any clear stakeholder analysis, either to identify specific users' level of interest and influence or to target the recommendations.

The accountability or organisational-knowledge purposes of the HIC and MSF(H) Darfur evaluations lent themselves to use at the organisational rather than field level. Although this meant that, for example, the MSF(H) field staff (rightly) viewed the evaluation as HQ-driven and (perhaps wrongly) as of little value to themselves, at least the purpose was clear. The approach (external evaluators, the buy-in of headquarters rather than field staff) largely matched the intended use of the evaluations.

According to the questionnaire and interviews for this chapter, it is relatively rare to design an evaluation with flexibility built into the ToR and budget, permitting adaptation in response to the context and emerging needs. While evaluators may voluntarily extend the time allocated to their work, this is more likely to occur before or after field trips. Matching the availability of multiple stakeholders with that of external consultants is extremely difficult. The interagency evaluation went to considerable lengths to be flexible but was constrained by the availability of staff in the field. Greater flexibility in the field would have to be planned in advance and built into the design and budget; for large-scale, real-time evaluations, this implies almost full-time availability.

Participation and ownership

The hypothesis of quality participation is that it will increase ownership and relevance. This is particularly important if the purpose is to generate learning. As noted in Section 3.3, crossing departmental boundaries to generate organisation-wide learning is vital but difficult and unusual. CAFOD's approach (Box 3.11) sought to do just this.

For Darfur, MSF(H) developed the ToR in Amsterdam, following a common process of then sharing it with the field and responding to suggestions. According to the evaluators, ownership was low in the field. There was no contact between the evaluators and the field before or after the evaluation. Had there been, despite the fact that the evaluation was delayed and the programme completed, there may have been opportunities to include objectives valued by the field staff (such as to contribute to new programme planning).

The interagency real-time evaluation (RTE) in Darfur was commissioned by the ERC/USG, and managed by OCHA. It required rapid deployment in order to be real-time. As a result the ERC/USG requested that OCHA HQ draft the ToR, which were circulated to IASC principals and then finalised without real field participation.

Box 3.11	Generating organisation-wide ownership

The organisation-wide ambitions of the CAFOD review (O'Donoghue, 2005) meant that the product had to be informed and owned by all of the senior managers. Established intra-organisational teams already in place provided an appropriate structure for organisation-wide engagement. The process adopted was highly participatory and iterative; a series of discussions took place and the ToR were adapted to evolving needs. The evaluator facilitated a series of group discussions, rather than concentrating on one-to-one interviews.

The effect of this approach was to capture the team and organisational dynamics, essential to a review pitched at an organisational level. This process of sometimes-intense debate meant that the recommendations evolved through an interactive series of edits, distillations and revisions. While the evaluator proposed options, the process itself meant that staff was engaged, 'translating' the findings, making them relevant to CAFOD and, most of all, owned.

The field viewed the evaluation as a donor-driven accountability exercise. Coming at a time when the response was scaling up in an intensely difficult context, the evaluation team was not, by and large, welcome. To make matters worse, agencies were suffering from a major shortfall in personnel; they 'needed staff, not scrutiny'.

This experience demonstrates the real challenge to an RTE and in particular to interagency RTEs. Rapid initiation is essential, but so is the full engagement of field personnel, given that they are the primary users. Had the ToR been designed in the field, the number of stakeholders involved and the complexity of achieving consensus would have severely delayed implementation. In fact it may not have happened at all; over-stretched personnel may not have elected to conduct an evaluation at that time.

A tentative conclusion is that an interagency RTE is an incompatible composite of evaluation types. In the future, a careful analysis of the trade-offs between timeliness and participation, as well as a ruthless prioritisation of purpose should underpin the design. The result may be a departure from the approach expected by some stakeholders and 'normal' evaluation practice. The first visit, for example, could be entirely dedicated to designing the evaluation, through a collaborative process involving the key field stakeholders and the evaluators. The result – in theory – would better suit the field users, but may not suit the information and accountability needs of others.

Planning

Most of the evaluators interviewed said that lead times for evaluations were often 'ludicrous' with far too little time to develop the process (an observation corroborated by the number of evaluations advertised through ALNAP which are expected to start almost immediately). Stories were told of last- minute contracts, ToRs being read on the plane, minimal contact with staff and few opportunities to establish and build relationships.

Similarly, many said they spend more time on an evaluation than had been allocated in the budget, indicating a mismatch between the planning and the reality – a perception not shared by the evaluation managers, who were more likely to think the time allocated was adequate. It may be that the managers are unaware of the time constraints experienced by evaluators. Either way, tight schedules reduce flexibility and ad hoc opportunities for learning.

An RTE also faces the particular challenge of taking place during particularly intensive periods of an emergency. The interagency evaluation team arrived in Darfur when agencies were intensively scaling up their activities. OCHA had also had 175 visiting delegations over the preceding 18 months.

Timing was an issue for both of the Darfur case studies. The MSF evaluation was delayed and took place too late to influence the programme, which was then approaching closure.

Last-minute planning, difficulties in recruiting evaluators and failing to allocate sufficient time may be more significant for low utilisation than is recognised. There is little recorded about such administrative constraints, and perhaps more attention should be paid to their impact.

The evidence

How the report is written is, according to interview findings, an important factor in take-up. The kind of social-sciences language sometimes found in evaluations is 'ghastly' and apparently reduces the readability of the report. Simple, concise and clear language was preferred. Aspects of language can be organisation-specific. MSF(H) for example dislikes academic or 'woolly' language which is ill-suited to MSF's straight-talking culture.

Most of all, interview comments related to the utility of the reports were linked to the length of the reports (too long and therefore not read) and the nature of the recommendations. Information overload is an increasing problem, and many personnel interviewed viewed the length of reports as an inhibiting factor. Recommendations are said to be too numerous, forming 'shopping lists' that are not prioritised or linked to identifiable individuals or even departments. Broad, strategic recommendations are difficult to action, involve multiple stakeholders and accountabilities and are less likely to be acted upon.

The extent to which recommendations should be generated through a collaborative process to ensure their relevance and feasibility is an interesting consideration for utilisation. This approach does not appear to be the norm. Interviewees suggested that it was common for the evaluation users simply to correct factual errors in a report. This is intended to preserve independence but ignores the fact that recommendations represent a process of operationalising the findings. Some agencies point out that external evaluators do not know enough about the organisation to do this. Others involved the evaluators as facilitators or combined internal and external evaluators. One of the case studies periodically referred back to the evaluator to ensure that the findings and spirit of the recommendations had

not been distorted or co-opted in the process, an interesting example of maintaining a balance between the independence and integrity of the evaluation and relevance to the organisation.

Follow-up mechanisms

Three of the four case studies used a matrix approach, with recommendations presented alongside actions, individuals or departments responsible, sometimes with a timeframe. One agency designed a clear mechanism for follow-up that was based around an individual focal point and was highly interactive. The focal point maintained momentum and played an active dissemination role – including the anticipation of concerns, additional information needs and potential difficulties before they could lead to blockages.

There was little evidence in our sample to demonstrate whether matrices and management plans of action promoted use or not. They certainly capture the findings and may be the only practical solution to following up on large evaluations. They can however become unwieldy. It is impractical to list individual responsibilities or posts if there are scores of recommendations and stakeholders. They cannot capture the complexities of implementing certain recommendations. Neither can they substitute for active management and accountability at all levels in an organisation.

> The ERC, OCHA and the IASC did not manage to get the full cooperation of the field for the exercise, but also did not insist on a more rigorous response and action from the field on the evaluation team's findings... Follow-up was constrained by several real and practical challenges, but it was also made possible by the lack of follow-up and ownership at the senior headquarters' level'. (OCHA, 2006, p3)

If action plans are not combined with a face-to-face approach, they can also invite an unproductive series of claim and counter-claim in which the user refutes a recommendation, the evaluation team responds, the user counters and so on. OCHA, initially relying on a matrix approach, subsequently recruited an experienced M&E specialist in a full-time follow-up post, based in Sudan. Follow-up took on a more strategic focus on systemic issues and worked on translating the recommendations that were still relevant into a dynamic management tool. OCHA

HQ personnel consider that this is an essential role; future evaluations will aim to have such a focal point in place earlier, facilitating and acting as a 'champion' for follow-up.

The sector is still learning how to establish meaningful and dynamic mechanisms for follow-up. The low involvement of the evaluators in this phase is striking; it may be pragmatic but surely undermines their ability to develop their craft (Box 3.12). The survey questionnaire showed that half the evaluators' clients rarely or never established follow-up mechanisms, giving an overall picture of weak or, at best, patchy follow-up planning. There is also little evidence of adapted and creative dissemination methods designed to suit different stakeholders; the norm continues to be a standard report and executive summary. This may indicate that the sector still has much work to do before the routine inclusion of follow-up mechanisms in evaluation planning is common practice.

The evaluators

The competence of the evaluators was not generally raised as an issue in the case studies. Credibility partly relates to the users. For example the studies indicate that field staff are more likely to view an evaluator as credible if they have an operational background and know the context (and preferably the organisation) well. A good skill match, such as skills in the management of organisational change for an evaluation leading to organisational learning, is important.

Box 3.12	Evaluators and follow-up

The lack of the evaluators' involvement following evaluations raises issues about their learning. Survey responses indicated that the provision of systematic and detailed feedback on the quality of the evaluators' work was patchy. None of the evaluators interviewed for the case studies had any idea of what had been done with their work or had any detailed feedback on the quality. Given the concerns about the quality of recommendations and the accessibility of formats and language, this is a missed opportunity for evaluators to increase their skills and the utilisation of their work. More involvement and feedback would also increase the ownership and accountability of the evaluators for the utility of their work.

Summing up quality factors

Evaluations that identify, prioritise and focus on the main purpose and primary users, designing the approach and product with their needs in mind, are more likely to be used. Planning for follow-up and making adequate time for meaningful participation and timely evaluation are essential to ensure relevance, ownership and systematic consideration of findings.

However, findings from research carried out for this chapter nonetheless suggest that the application of such pro-use approaches is patchy at best. In addition to relational and organisational factors, it is possible, though not at all proven, that this is due to limited awareness of quality factors that can enhance utilisation. There is quite a body of information available, but it is dispersed among numerous individual agency studies and research papers that many practitioners may not have seen. One key reference, *Utilization-Focused Evaluation* (Patton, 1997) runs to 400 pages of close print; not all will or can find the time to read – and then adapt – such information. There could be value in summarising the practices in a brief guideline or checklist format that is more accessible and easier to operationalise.

3.4.3   Relational factors affecting use

The second main group of factors affecting utilisation is based around links and relationships between the evaluation community and the users. MSF(H)'s nutritional policy was influenced by participation in networks and by its peers in the MSF Movement, themselves influenced by the community-based approach seeded in the Bahr el Gazal evaluation. In a sense, MSF(H)'s participation in sectoral and peer networks generated a boomerang effect, in which the influence of the evaluation returned to it via the networks and peers.

The facilitation and interpersonal skills of evaluators are an important influence. The ability of evaluators to create a learning atmosphere is important, if that is the purpose of the evaluation. It can involve skilfully managing the border between what remains unsaid and what becomes black and white in the report. The attitude of the organisation and its staff to learning and accountability will also affect the ability of an evaluator to exploit opportunities for active learning. In one of the case studies, the evaluator had worked for the agency previously; he was known and trusted. In other cases, the relationships were variable. On occasion, the frustration of field staff

with an evaluation was transferred to the evaluators. The personalities involved can be a significant factor, limiting opportunities for the evaluators to build relationships and undermining use from the outset.

There is a tendency to expect staff members to accept criticism gracefully and constructively. In reality, this often does not happen. The interpersonal and conflict-management skills needed to manage defensiveness and opposition are rarely mentioned as essential competencies. In general, the interpersonal dimension of the evaluation process goes unrecorded, despite the fact that it may be a significant factor in the quality and utilisation of an evaluation.

It is worth noting that evaluators were regarded with considerable suspicion by some of the senior and field personnel interviewed. The evaluator may be regarded as a harbinger of doom, or a nuisance and diversion from work. A common perception is that evaluators are too concerned with their reputations and less concerned about the utility of their work. Interviewees complained that evaluators did not recognise the provenance of many of the solutions proposed, which may have been identified by staff or already been implemented. Evaluators are sometimes viewed as unwilling to make information more user-friendly (by avoiding evaluation jargon) on the basis that their reputations would be damaged if they did not comply with the expected standards of a good evaluation report.

The commercial nature of the relationship with external evaluators can also be an issue, particularly in smaller organisations and for NGO field personnel who are more likely to regard consultancy fees as excessive and the work as mercenary. Some evaluators interviewed noted a tendency of clients or field staff to foist an almost inhuman schedule on the team, as if the most had to be obtained from every invoiced day.

Overall, insufficient attention is paid to the relational side of evaluation. Using evaluation is as much a people issue as it is a technical one – and perhaps more so. The findings of this study suggest that evaluators and managers need to invest more to generate trust and bridge the gap between users and evaluators. If the focus of the evaluation is on learning, then the evaluators and users should be in partnership, pooling experience and skills in the service of improved performance.

3.4.4 Organisational factors affecting use

Examining the learning cultures, structures and knowledge management of the case-study organisations was beyond the scope of this study. Of interest is the effect of management and political issues on their utilisation of the evaluations. None of the agencies involved, whether they have relatively hierarchical or 'flat' organisational structures, can simply communicate directives to change policy or practice and expect them to be implemented. Often, whether or not policy is changed in response to evaluations is dependent on the convictions of individual stakeholders. Decentralised agencies have highly autonomous field decision-making and the links between the programme and the headquarters sectoral or policy staff are relatively weak; headquarters personnel are largely advisory. Translating policy into action in the field can be difficult.

The relationship between the evaluation and decision-makers for the interagency Darfur evaluation could hardly be more complex. The ERC/USG commissioned the evaluation but does not 'line manage' the heads of the UN agencies in Khartoum. The Humanitarian Coordinator's authority over agencies in the field can also be contested, particularly by the NGOs. An interagency Core Learning Group (CLG) was established at HQ level, but was not populated by decision-makers. The Steering Committee for Humanitarian Response, representing the NGOs on the CLG, had no authority over the operational decision-making of its members in Darfur. The lessons-learned paper prepared by OCHA (2006) notes that 'a great deal of the lack of ownership can be attributed to institutional insularity, which led to a lack of accountability for the exercise at the individual agency, as well as collective IASC, level' (p4).

Concerns for an organisation's reputation featured in some of the case studies. In Darfur, agencies were concerned that 'highlighting and publicizing gaps in the response would dash donor confidence and thus discourage support at a time when low funding was seen as a key obstacle to improving the response' (OCHA, 2006, p3).

The sense of an evaluation being imposed was a clear issue, although in these examples it was usually an internal issue between headquarters and the field, rather than a reflection of unequal aid relationships between organisations. The electronic

questionnaire conducted for this study asked respondents what, in their view, was the most common reason for commissioning an evaluation, choosing between four categories of formal accountability (it has to be done, rarely commissioned by the organisation), improving the programme, organisational learning and legitimising. For the majority of the evaluators, formal accountability was perceived to be the primary motivation for commissioning an evaluation. This may indicate that 'ritual' evaluations are common.

The top-down imposition of an evaluation was cited as one of the three main factors inhibiting use by half the evaluation managers and over a third of the evaluators responding to the survey. If ritual and imposed evaluations are indeed common, then the potential for utilisation is greatly reduced even before the evaluation has begun.

MSF(H) and CAFOD have both taken proactive and robust approaches with donor-commissioned evaluations to ensure that the evaluation was of value to both partners. CAFOD and MSF are unusual in that a high percentage of their income is private. Other agencies appear less confident in regard to influencing the purpose, approach and use of evaluations they have not themselves commissioned. Good practice by some donors is in evidence through the delegation of evaluation to their funded partners, an approach largely adopted by USAID. ECHO is in the process of developing an evaluation methodology that can be managed and implemented by the partner NGO itself. The 'push' is still there, but the partner can design and manage the evaluation to optimise learning.

Stakeholders, including the evaluators, have the right to question intended use. The fear of losing funding or of appearing to be resistant to learning makes agencies timid about challenging the purpose of an evaluation and demanding greater partnership in the process. Similarly, all stakeholders have a duty to consider whether the financial and opportunity costs of evaluation can justify a symbolic or administrative need. Something cheaper and quicker may suffice.

3.4.5 External factors affecting use

This study has found little evidence of external pressure acting as a force for the utilisation of evaluations. In fact, it is noticeable by its absence. The questionnaire

responses showed that it was unusual for the key findings of an evaluation to be communicated to the programme's beneficiaries in the field. No cases were found in which an NGO, for example, held public feedback sessions with community leaders.

Most of those interviewed said that evaluation was important, but was only one of the resources and influences for change. It was generally given a middle ranking in terms of its value to decision-makers. Influential factors appear to vary, in part according to the particular focus or culture of the organisation. Some are strongly influenced by peers, others by technical reviews, visiting experts and participation in networks related to particular sectors. Senior managers often referred to the importance of discussion with staff. The use of a variety of sources of learning is of course positive. But it implies that evaluation planners and evaluators could do more to recognise the relationship between an evaluation and other learning processes. Playing a modest role does not undermine the value of an evaluation. There are other factors that do.

Most of the literature looks at change in only one direction – from the evaluation forwards to its utilisation. Although questions were included in this study regarding other influences on learning, the scope precludes any conclusion. Responses however indicate that most people and organisations do not seem to have given this a great deal of thought. What has worked? What are the most powerful agents of change in the history of an organisation, or in the sector? A rich line of enquiry could be to trace backwards from identifiable changes and explore the influences and pressures that led to it. A better understanding of evaluation's role in this process could enhance its design, or perhaps lead to the invention of different tools altogether.

3.4.6   Limitations to the utilisation of evaluation

Utilisation is only partly controlled by evaluation quality and process. Its ability to effect change in the context of powerful organisational obstacles and political imperatives is relatively modest. There are many competitors for an organisation's attention, and evaluation often loses out to other priorities. Evaluation is one (competing) part of a decision-making process. 'The policy-making process is a political process, with the basic aim of reconciling interests in order to negotiate a consensus, not of implementing logic and truth' (Weiss, 1977, p533).

The entirety of the utilisation process is not under the control of any one person or process. The more complex the use, the more this is so. Arguably, the evaluation is but a catalyst, influencing the beginning of a process and increasingly ceding control to the impact of other actors and other forces. Humanitarian workers have the responsibility to make sure that the inputs, activities and outputs of a project are the best possible basis for the intended outcome and impact of the programme, but they cannot predict exactly what those higher-level results will be. The influence of an evaluation similarly reduces in inverse proportion to its appropriation by the users (Figure 3.2).

Figure 3.1 The decreasing influence of evaluators in the change process
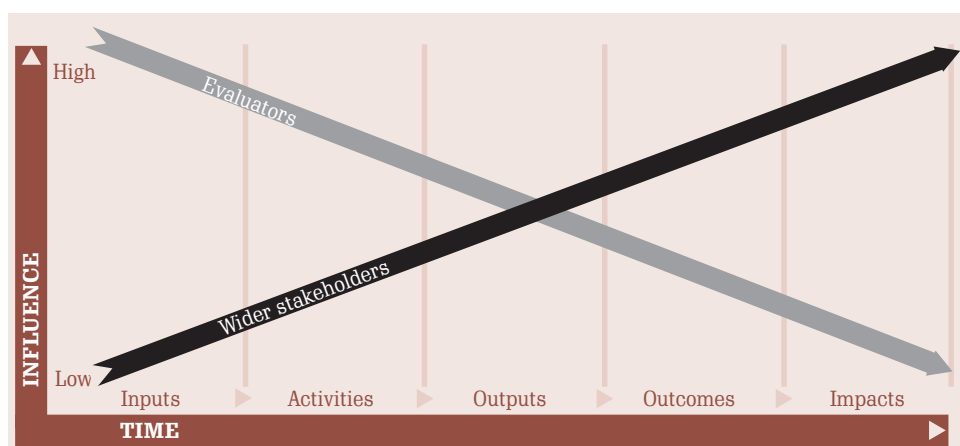
(Adapted from Ramalingam, 2006)

The core responsibility of the evaluators and managers is therefore to provide a good foundation for use. After that, the use is in the hands of the users – and so it should be. Although there may be an argument for the evaluators subsequently to play an advisory role (exemplified in the follow-up to the JEEAR), neither evaluator nor evaluation can be the sole agent of change. In addition, there are limits to the change that the sector can bring about alone.

Some of the angst surrounding non-utilisation may be misplaced. In part, it is the 'recurring structural impediments' (ALNAP, 2005, p25) that lead to the conclusion that evaluation is failing to contribute to improved performance. In part, this is true. At the same time, many persistent weaknesses relate to structural characteristics of the humanitarian sector that aid agencies argue are beyond their control or mandate.

The main recurring themes identified each year by the ALNAP synthesis are: human-resource problems (such as lack of training and trained staff, high turnover); competitiveness and poor coordination; weak links between relief, rehabilitation and development; the politicisation of aid and capacity building of national partners. Many of these weaknesses are at least partly the consequence of unpredictable funding and the earmarking of income for operational response rather than institutional development.

As long as aid is linked to government money and the public's expectations of charity in the north – which still expects white faces, quick solutions and good visuals to illustrate good performance – there are limits to the sector's latitude for change. Evaluation recommendations that, for example, propose reduced staff turnover and longer contracts may not be implemented because it is simply impossible for the agency to find, retain and fund those personnel.

Conversely, structural issues can be used as an excuse for inaction and conceal an unwillingness to challenge the status quo and seek innovative solutions. Should an evaluation concentrate only on what is within the control of an agency to change or insist on reference to an aspirational yardstick? Comparison with an ideal performance will find all humanitarian response wanting. There is an argument for evaluators at least to differentiate between what the potential users could do, and what should be done if other conditions were met. Evaluation is then fully contextualised at all levels.

This would again increase the need to target the relevant decision-makers. The decision-makers who can influence the systemic issues undermining performance are not typically the audiences of evaluations. USAID may seek to increase funding for Darfur but it is not the US Congress. Evaluation appears, at any rate, to be poorly suited to influencing political agendas. How much effort is made to ensure that the system-wide problems are 'translated' and communicated for the consumption of the right audience, or are we just talking to ourselves?

At present, evaluation is expected to serve both the day-to-day and loftier objectives of the sector. As well as serving learning needs, administrative formalities and information provision, there seems to be an expectation that evaluation will also generate organisational restructuring, strategic reorientation and influence on the wider political landscape in which the sector operates. This suggests either that evaluation as a tool has to adapt considerably to suit these different objectives, or

that broad recognition is needed that evaluation makes a valuable if modest contribution to a menu of tools and strategies for change.

Evaluation viewed as one of several tools at once enhances its potential to influence and demotes it from its position as a catch-all for accountability and learning. As other studies have noted (eg Ramalingam, 2005), learning strategies work best when they are an integrated component of a dynamic, phased and progressive approach to organisational development. Perhaps this, rather than the isolated utility of evaluation, should be the primary concern and focus of future work on learning and accountability.

3.5 Conclusion and ways forward

3.5.1 Concluding remarks

This chapter began with a concern that the credibility of evaluation will be undermined if its poor record of influencing humanitarian performance continues. In some ways then, the findings indicate a less gloomy outlook: our pessimism partly results from a narrow perception of utilisation that does not do justice to the rich and often indirect use and influence of evaluation.

At the same time, this picture of utilisation makes it difficult to assess the effectiveness of evaluation itself. Its utilisation is largely a mirror of the complex nature of change; it is transformed by the users themselves, strongly affected by context and often unpredictable. Evaluation is also irrevocably bound up with the learning culture and absorption capacity of the organisations and individuals involved. 'Serious attention to use involves financial and time costs that are far from trivial' (Patton, 1997, p383). A user-based approach is also challenged by the sheer number and geographic distribution of stakeholders that characterise humanitarian response. This means that even if the evaluation community gets better at applying a utilisation-focused approach (as it clearly should), there are real limits to how often such evaluations should be conducted and to the changes that evaluation can be

expected to lead to. This is true of any single learning and accountability mechanism that must also compete for attention with a diverse and often political range of organisational pressures and individual learning preferences.

If we continue to expect evaluation to cover most of the accountability and learning needs of the sector, we will be disappointed. It is critical that the job expected of evaluation is realistic, clear, better differentiated from other objectives and designed accordingly. Evaluation is often squeezed into multiple roles for which it is not always best suited, or that contradict each other. It may not necessarily be the most cost-effective approach for certain uses. It is not, despite its reputation as a tool for upwards accountability, widely used to hold organisations to account. Learning is highly dependent on genuine participation – which is not necessarily a feature of many of the time-poor evaluations motivated by administrative ritual and symbolic accountability or characterised by the report and adherence to predetermined methods. The use of many evaluations is confined to particular groups, often excluding the very stakeholders who are key to cross-departmental decision-making, and providing an unlikely basis for organisational learning.

The responsibility of evaluation lies in identifying the strengths and weaknesses of programmes, which it can do extremely well, and in facilitating utilisation, which it has been doing less well. Serious participation and a far greater focus on the intended user and uses would help to expose the practice of inappropriate or ritual evaluation and prevent evaluation further contributing to the current mistrust and saturation in the sector.

Much more effort may also be needed to identify better organisational-change agents to continue the process that accountability tools can start but not complete. If external pressure is a significant change agent, then audits may serve upwards accountability better than evaluations, linked to regulatory bodies designed for the purpose. Participatory and impact evaluation designed with, and for, the affected communities would strengthen downwards accountability. Organisational and individual learning is better served by interactive approaches characterised by openness, not judgement. Isolated efforts to innovate, such as those implemented by ActionAid, can be built upon and shared to make learning and accountability experimental and creative, rather than potentially moribund and mistrusted.

3.5.2 Considerations for the future

The focus of accountability tools should shift from methodology to utilisation. If the goal is to enhance performance, then the door is open to a range of possible tools, of which evaluation is only one. To an extent, different expectations of evaluation derive from the belief system that underpins accountability. Some users and evaluators perceive a wilfully recalcitrant sector that must be forced to change. This view will increasingly lead to an increase in tools and evaluations to measure performance, compliance with standards, accreditation and so on. This is similar to the route already taken by the public sector in many developed countries. Evaluations serving this purpose would be predominantly formal and published. Consideration would have to be given to ensuring that such evaluation teams were given the kind of mandatory access and cooperation that is commonly the preserve of financial audits.

Others view the humanitarian sector as, essentially, populated by highly motivated personnel doing their best in difficult circumstances. This latter viewpoint lends itself to an approach to evaluation that is predicated on trust and the creation of safe spaces for learning, including making mistakes. Either way, the choice of tools that best generate change would be based upon a clear analysis of which approach works best. Some organisations may respond better to the stick and others to the carrot. The impact of evaluation is enhanced, if not enabled, by being part of a broader menu of approaches to enhancing performance. Monitoring, for example, remains a poor cousin of evaluation and has yet to receive the same attention from decision-makers. Evaluation, in whatever form, is only one element of accountability.

If the indications in this study, that evaluation has a relatively poor image with operational staff and senior managers, hold true more broadly, then there is work to be done to reduce the gap between evaluation and users. A weakness of this study was limited consultation with field personnel and senior managers. This is also a weakness of other studies. Given that these groups are, or should be, essential users of evaluation, this is a curious oversight. It would be useful to pursue this line of enquiry, finding out what kind of accountability and learning tools such users would value.

It may be that much more could be done to adapt evaluation to serve organisational learning better. However, evaluation may not be the best tool for this at all; more

appropriate methods may be found in approaches currently used in the private and public sectors. More attention could be paid to what has actually led to enhanced performance, both within specific organisations and across the sector, by examining significant changes in performance and tracing backwards to the causes.

Whatever the accountability tool, no instrument will generate change unless the sector finds the time and space to allow it. At present, there is a sense that the humanitarian sector is suffering from inertia generated by an overload of tools, guidelines and bureaucracy. Evaluation itself may be inadvertently contributing to the workload. The decision to carry out an exercise in learning and accountability should be carefully considered; more robust prioritisation of evaluation by all stakeholders may be necessary now. Part of the menu of tools could simply entail giving staff the time to innovate and absorb information. Perhaps there is just one leaf to take out of Google's book, which is reported to include giving 'each employee a day a week to work on their own projects, the genesis of many of the new services the company is rolling out every other week' (*Observer*, 3 September, 2006).

References

ALNAP (1997) *Organisational and Institutional Learning in the Humanitarian Sector. Opening the Dialogue*. Koenraad van Brabant. London: ODI.

ALNAP (2002) *Humanitarian Action: Improving Performance through Improved Learning*. London: ODI.

ALNAP (2003) *Annual Review 2002*. London: ODI (http://www.alnap.org/AR2002/ AR2002_fulltext.pdf).

ALNAP (2004) *Review of Humanitarian Action in 2003. Field Level Learning.* London: ALNAP (http://www.alnap.org/RHA2004/ index.html).

ALNAP (2005) *ALNAP Review of Humanitarian Action in 2004. Capacity Building.* London: ODI.

Broughton, B and S Maguire (2005) *Inter-agency Real-time Evaluation of the Humanitarian Response to the Darfur Crisis*. OCHA.

Buchanan-Smith, Margie (2005) 'How the Sphere Project Came into Being: A Case Study of Policy Making in the Humanitarian-aid Sector and the Relative Influence of Research' in J Court, I Hovland and J Young (eds) *Bridging Research and Policy in Development. Evidence and the Change Process*. London: ODI (http://www.odi.org.uk/ rapid/Publications/BRP_ITDG.html).

Carlsson, C and P Engel (2002) *Enhancing Learning through Evaluation: Approaches, Dilemmas and Some Possible Ways Forward* (background papers). Maastricht: ECDPM (http://www.ecdpm.org/ Web_ECDPM/Web/Content/ Navigation.nsf/index2?readform&http:// www.ecdpm.org/Web_ECDPM/Web/ Content/Content.nsf/0/ 05A45EF48F023482C1256C7E00555A9C).

Carlsson, J, M Eriksson-Baaz, AM Fallenius and E Loevgren (1999) *Are Evaluations Useful? Cases from Swedish Development Co-operation.* Sida Studies in Evaluation (http:// www.sida.se/sida/jsp/ sida.jsp?d=118&a=2355&language=en_US).

CGD (Center for Global Development) (2006) *When Will We Ever Learn? Improving Lives through Impact Evaluation. Report of the Evaluation Gap Working Group*. May. Copyright © 2006 by the Center for Global Development, ISBN 1-933286-11-3.

Charity Commission (2005) *Report of Findings of a Survey of Public Trust and Confidence in Charities*. Prepared by Opinion Leader Research for The Charity Commission, November.

Church, Cheyanne (2005) *Mind the Gap: Policy Development and Research on Conflict Issues*. INCORE, January (www.incore.ulster.ac.uk).

Corbett, M and A Lloyd (2005) *MSF Holland's Medical and Nutritional Strategy in Darfur – Sudan.* MSF, May.

Edgren, Gus (1999) *Fashions, Myths and Delusion: Obstacles to Organisational Learning in Aid Agencies* In Jerker Carlsson and Lennart Wohlgemuth (eds) *Learning in Development Co-operation*. Stockholm: Almqvist and WIksell International.

Fritz Institute (2006) *Evidence of Impact: Challenges and New Directions*. Impact Conference, 19–20 May, 2006, Sebastopol, California Conference Proceedings.

Gordillo, G and K Anderrson (2004) 'From Policy Lessons to Policy Actions: Motivation to Take Evaluation Seriously' in *Public Administration and Development* 24, pp 305–320 (http://www3. interscience.wiley.com/cgibin/abstract/ 109629683/ABSTRACT?CRETRY= 1&SRETRY=0).

HAPI (2006) *The Accountability Alibi*. N Stockton. Humanitarian Accountability Partnership International, May (www.hapinternational.org).

HPG (2002) *Financing International Humanitarian Action: A Review of Key Trends*. HPG Briefing Paper, ODI, November.

Kent, R (2004) *Humanitarian Futures. Practical Policy Perspectives*. HPN paper number 46, April. London: ODI/ Humanitarian Practice Network.

Landry, Réjean, Moktar Lamari and Nabil Amara (2001) *Extent and Determinants of Utilization of University Research in Public Administration*, 29 July to 1 August, Canada (http://www.fss.ulaval.ca/kuuc).

Mango (2004) *Helping People is Difficult: A Contribution to Recognising the Constraints and Releasing the Energies of NGO Staff*. A Jacobs, 18 November.

MFA Netherlands (2000) *Evaluation Feedback: Experiences and Recent Changes*. R. van den Berg and T Kliest. Policy and Operations Evaluation Department, Ministry of Foreign Affairs of the Netherlands. Paper prepared for the OECD/DAC Working Party on Aid Evaluation Workshop on Evaluation Feedback for Effective Learning & Accountability, Tokyo, 26–28 September 2000.

MFA Finland (1997) *Utilisation of Evaluation Results in Finlands' International Development Cooperation*. Ministry of Foreign Affairs, Department for International Development Cooperation, Finland, June.

NAO (2005) *Report by the National Audit Office: Engaging with Multilaterals*. National Audit Office, Department for International Development, December.

OCHA (2006) *Lessons Learned in Conducting the Inter-Agency Real-Time Evaluation of the Humanitarian Response to the Darfur Crisis*. Prepared by OCHA's Evaluation and Studies Unit (OCHA/ESU), January.

O'Donoghue, G (2005) *Rudder and Sails. A Review of CAFOD'S Humanitarian Function*. CAFOD, June.

OECD (2000) *Evaluation Feedback: Experiences and Recent Changes*. R Van den Berg and T Kliest, Policy and Operations Evaluation Department, Ministry of Foreign Affairs of the Netherlands. Paper prepared for The OECD/DAC Working Party on Aid Evaluation.

OECD (2001) *Evaluation Feedback for Effective Learning and Accountability* (http://www.oecd.org/dataoecd/10/29/2667326.pdf).

O'Reilly, F and J Shoham (2002) *MSF Holland in Afghanistan Mission Evaluation: May 2000 – May 2002*. MSF.

Patton, M (1997) *Utilization-Focused Evaluation*. London: Sage.

van de Putte, B (2000) *The Utilisation of Evaluation Recommendations in Medecins Sans Frontieres – Holland: A Study of 10 Evaluation Reports (1997–1999)*. Amsterdam: MSFH, May.

van de Putte, B (2001) *Follow-up to Evaluations of Humanitarian Programmes*. London: ALNAP, April (http://www.alnap.org/pubs/pdfs/followup.pdf).

Ramalingam, B (2005) *Implementing Knowledge Strategies: Lessons from International Development Agencies*. Working Paper 244, ODI, April.

Ramalingam, B (2006) *How Can Outcome Mapping Help Strengthen Knowledge and Learning Strategies?* Presentation given at Crossing Borders, Crossing Boundaries, 2005 Joint AEA/CES Conference, 24 to 29 October 2005 (downloaded from http://www.idrc.ca/en/ev-91687-201-1-DO_TOPIC.html).

Ridde, Valéry and Shakir Sahibullah (2005) 'Evaluation Capacity Building and Humanitarian Organization' in *Journal of Multi Disciplinary Evaluation (JMDE)*, 3 October.

Sida, L and Chris Szpak (2004) *An Evaluation of Humanitarian Information Centers including Case Studies of HICs*

for Iraq, Afghanistan, and Liberia. USAID/ DFID.

Shulha, L, M Cousins and J Bradley (1997) 'Evaluation Use: Theory, Research and Practice since 1986' in *Evaluation Practice* 08861633, Vol 18, Issue 3, Fall.

Slim, H (2002) 'By What Authority? The Legitimacy and Accountability of Non-governmental Organisations' in *The Journal of Humanitarian Assistance,* 10 March (http://www.jha.ac/articles/ a082.htm).

Suhrke, A (1999) 'From One Crisis to Another: Organisational Learning in UNHCR' in Jerker Carlsson and Lennart Wohlgemuth (eds) *Learning in Development Co-operation*. Stockholm: Almqvist and WIksell International.

Telford, J and J Cosgrave (2006) *Joint Evaluation of the International Response to the Indian Ocean Tsunami: Synthesis Report*. London: Tsunami Evaluation Coalition.

UNHCR (2001) *Real-time evaluation of UNHCR's response to the Afghanistan emergency*. Bulletin No. 2, J Crisp and E Stiger.

Valid International (1999) *MSF in Catastrophe*. Steve Collins, Valid International.

Weiss, Carol (1977) 'Research for Policy's Sake: The Enlightenment Function of Social Research' in *Policy Analysis* Vol 3, No 4, pp 531–545.

Wigley, Barb (2005) *The State of UNHCR's Organization Culture*. EPAU, UNHCR.

Wilding, J, Issaka Idrissa Mossi, Debra Lynne Edwards, Amanda Weisbaum, Boube Aw, Timothy Mander (2005) *Joint Independent Evaluation of the Humanitarian Response of CARE, CRS, Save the Children and World Vision to the 2005 Food Crisis in the Republic of Niger.*

Williams, K, B de Laat and E Stern (2002) *The Use of Evaluation in the Commission Services – Final Report* (http:// europa.eu.int/comm/budget/evaluation/ pdf/use_of%20_evaluation_final_ report.pdf).

WFP (World Food Programme) (2005) *Study on WFP Follow-up Evaluation Recommendations*. Executive Board Summary (Document Number: WFP/ EB.A/2005/7-B) (http:// documents.wfp.org/stellent/groups/ public/documents/eb/wfp050928.pdf).

Wright, Malaika and Pauline Wilson (2006) 'Can Joint Evaluations Promote Ongoing Collaborative Action by NGOs? (Emergency Capacity Building Project)' in *Humanitarian Exchange* No 34, June, ODI/ Humanitarian Practice Network.