

Improving humanitarian
impact assessment:
bridging theory and practice

Karen Proudlock and Ben Ramalingam
with Peta Sandison

The focus of humanitarian evaluations is shifting towards analysis of the impact of humanitarian assistance – to understand, in an evidence-based way, how aid ultimately affects the lives and livelihoods of aid recipients.

Improved impact assessment could contribute to greatly improved beneficiary participation, more robust needs assessments, and more evidence of what works.

This review summarises current debates on impact assessment in humanitarian aid, and identifies a series of challenges. It is based on research including a literature review, debates with ALNAP members, interviews with key informants, and detailed case studies of four humanitarian initiatives on impact assessment.

Key message 1

Humanitarian impact assessment focuses on the effects of humanitarian aid, and reflects increasing concern to better understand what aid achieves. In order to be credible, impact assessment requires greater involvement from beneficiary populations and local actors. Humanitarian impact assessment is still relatively unusual, often being perceived as too difficult and too expensive. However, evidence from impact assessment efforts suggests that there are ways to navigate these challenges.

In a context of limited resources and increasing debate about humanitarian performance, there is now considerably more analysis and scrutiny of humanitarian operations. Evaluation of humanitarian action now attracts increasing levels of donor funding and agency commitment, as well as public and political interest. Yet, the role of beneficiaries remains peripheral or at best partial in the planning, monitoring and evaluation of humanitarian aid. Moreover, evidence about the impact of aid interventions is scarce, and impact assessment is held by some to be insufficiently rigorous in the sector. Despite this, evidence shows that humanitarian impact assessment is possible, and there are different ways to determine the causes of change in humanitarian conditions in different contexts.

Key message 2

Calls for more and better impact assessment must be supported by shared approaches to understand and navigate the challenges faced. Work done as part of this study shows that humanitarian impact assessment requires coordinated, sustained and joined up effort in five key areas. Together, these five areas form a suggested conceptual framework which can be used as an analytical tool and as a starting point for developing and improving practice.

There are five key areas to think about with regard to any impact assessment initiative. First, an impact assessment must balance the priorities and interests of a range of different stakeholders. Is the assessment for learning or accountability? Is it for donors, the agency, wider academic research, or the affected people? Second, how should humanitarian impact be defined – impact on what, and over what timescale? Third, how can impact be measured? What indicators are appropriate, and against what baselines or comparison groups? How can it be proved that any observed or reported effects are actually caused by a particular intervention? What methods are appropriate to the given context, and how will issues of data, baselines, and timing be addressed? Fourth, how should data on impact be analysed and interpreted, and what role should affected people play in this? Fifth, how can incentives and capacities be developed to enable and improve humanitarian impact assessment?

Key message 3

It is necessary to make difficult and political choices in determining the balance of interests and priorities in an impact assessment. There are many ways to approach impact assessment, and consequently great diversity in possible purpose, scope and design. Case studies show the importance of negotiating stakeholder ownership of impact evaluations as early as possible in the process of design and implementation.

Like all humanitarian evaluations, and indeed all approaches to performance, impact assessments are often determined by two institutional priorities: accountability and

learning. Yet, there is an inherent tension between these two objectives. For example, the framing of impact assessment in terms of accountability for results can encourage risk aversion among humanitarian agencies and thereby undermine the learning and innovation required for improving performance. The overall focus of an impact assessment may be influenced by numerous stakeholders including beneficiaries, donor governments and publics, recipient governments, national and international NGOs, the media and researchers. It may be that the interests and aims of all these groups are not reconcilable or achievable. However, evidence shows that the findings from impact assessments are more likely to be used if the design and process fits the needs and interests of the end users. And choices regarding the purpose and scope of impact assessments will shape the selection of methodologies, the kinds of knowledge and conclusions generated, and follow-up and use of these. It is therefore essential to allow adequate time for the meaningful participation of all relevant stakeholders in defining the purpose and scope of impact assessments.

Key message 4

To enable effective assessment of the impacts of a humanitarian intervention, its underlying logic or ‘theory of change’ must be clear and explicit, not over-ambitious and based on a solid understanding of humanitarian needs. Impact can be defined in various ways, but it is most important to clarify its meaning in ways that are specific to particular interventions or contexts, and that enable practical implementation of an effective assessment.

It is generally agreed that impacts can be positive or negative, primary or secondary, direct or indirect, and intended or unintended. Definitions useful in humanitarian impact assessments may also specify impacts as the changes in people’s lives attributable to an intervention. This definition emphasises the ultimate effects on individuals and communities – whether people are better off because they are more independent or safer, rather than whether certain amounts of aid have been distributed or camps established. Given the often immediate and rapid nature of humanitarian interventions, the impacts may not necessarily be long-term. Both the

importance of attributing changes to a particular source, and the complexity of doing this, are reflected in the emphasis on the possibility of negative and unintended impacts. This can be seen as the difference between ‘effectiveness’, which focuses on the intermediate objectives of an intervention, and ‘impact’ which goes further in looking at the intervention in the wider socioeconomic and political context. Any practical definition of impact will depend entirely on the specific context and the goals of the programme being assessed.

Key message 5

Impact assessment inevitably involves value judgements about which kinds of changes are significant, and for whom. The choice of indicators and methods is therefore critical, yet hard to determine in advance and out of context. There is a range of ways to generate useful measures of impact, and there are a range of methods that are appropriate to distinct contexts, needs and budgets. As cited later: ‘methodological appropriateness could be considered the “gold standard” for impact evaluation’.

Impact assessments should ideally be carried out when impacts are likely to be visible and measurable, and not finished or forgotten, so timing of the assessment is critical. Aid recipients should be involved in identifying and selecting indicators to measure impact, because their choices may differ from those of other stakeholders. While baseline data collected before an intervention are useful in measuring impact, they are not essential. Impact can still be assessed against perceptions of how things were before the intervention, or in comparison with otherwise similar groups who did not receive the intervention. Either qualitative or quantitative data can be used to assess impact, but the most effective assessments often involve a mixture of both approaches. They are often complementary and can in combination enhance confidence in identification of the causal impacts of humanitarian programmes. Rather than asking, ‘which methods are appropriate for assessing humanitarian impact?’, it is more useful to ask, ‘which methods are appropriate, feasible and worthwhile under these specific conditions?’

Key message 6

Impact assessment involves the analysis and determination of causality, to show that observed changes are caused by a particular programme or activity rather than by other factors. There are two broad approaches to this – comparative approaches, and theory-based approaches, and both have distinctive benefits and drawbacks in the context of humanitarian aid.

Establishing causality in humanitarian situations is complex because results are unlikely to be due to any single factor. There are two broad approaches to determining attribution in humanitarian impact assessment: comparative methods attempt to establish what would have happened without a particular intervention, and theory-based methods examine a particular case in depth to explain how an intervention could be responsible for specific changes. While there are understandable ethical reservations about withholding an intervention from certain groups for the purpose of assessing its effects, comparative methods can be useful in providing more information about the particular merits of different aid strategies, such as the choice between cash or food aid, or the benefits of innovations relative to existing practices. Theory-based methods require that causality must be inferred from information about the context, from beneficiaries and key informants and by triangulation with data from other sources. It is important to note that comparative and theory-based approaches are not mutually exclusive. For example, there is no reason why a randomised evaluation could not be embedded within theory-based approach. However, at present, most impact assessments do not combine the two approaches.

Key message 7

Crisis-affected people and local and national actors are often crucial to assessing impact, as they know the context best, and are best placed to assess how lives have changed as a result of aid.

Participation of crisis-affected populations is not yet a standard feature of evaluations, and when it does happen it is either informal and opportunistic, or formalised in the form of surveys. The process of impact assessment, and

particularly the analysis and interpretation of results, can be greatly improved by the participation of affected people, who are the best judges of their own situation. The process can also assist communities and NGOs to measure impact using their own indicators and their own methods, overcoming the weaknesses inherent in many donor and NGO evaluations which focus on process and delivery over results and impact. Evidence from case studies shows the benefits of partnerships between donors, implementing agencies, communities and other stakeholders – the latter should include national and local actors as far as possible.

Key message 8

Improving accountability for humanitarian impacts requires a cultural shift, and clear articulation of how agencies can benefit from assessing their impacts. Evidence suggests that there is still some way to go in this area, and there is a lack of organisational capacity and incentive for humanitarian agencies to carry out impact assessments. Investment in sustainable partnerships may be key to positive changes in this area.

Many commentators note a lack of individual and organisational capacity to carry out good impact assessment within the humanitarian sector, despite a wealth of tools and methods. High staff turnover, lack of a 'learning culture' and inadequate resources are contributing factors. Disincentives for effective evaluation, learning and transparency include the complexity of impact-assessment methodologies, the rigid nature of agency programme and budget cycles, and aversion to the perceived risks of failure. The development sector includes several system-wide initiatives to strengthen capacity in impact assessment, but there is no equivalent in the humanitarian sector. This potential is yet to be explored, and could include scope for addressing the associated disincentives and costs. Evidence suggests that the design and implementation of impact assessments requires skills available only through investment in long-term partnerships between academics, donors, governments, practitioners and targeted recipients. Such approaches to impact assessment will involve learning that leads to improvements and innovations in how humanitarians operate. Ensuring the institutional sustainability of impact assessment in this way will also help to mitigate the considerable costs of impact assessments.

Key message 9

The final point is a note of caution: the desire to prove impact is a complex one, with many different underlying motivations. The desire to prove impact, if implemented in a narrow and self-interested way, may lead to never in fact *improving the real impacts* of humanitarian work. There is therefore a crucial need to ensure that the ongoing impact-related debates, actions and practices keep at their core the notion of an approach to impact assessment that is shaped by humanitarian principles first and foremost.

The final point is a note of caution: the desire to prove impact is a complex one, and one with many different underlying motivations.

If, as impact assessment gains more prominence, attention and resources, it is pursued first and foremost for narrow institutional or political purposes, the likelihood is that it will become yet another instrument that has failed to deliver on its early promise to improve humanitarian performance. The desire to *prove impact*, if implemented in a narrow and self-interested way, may lead to never in fact *improving the real impacts* of humanitarian work.

There is therefore a need to ensure the ongoing debate on impact – and the actions and practices that result from it - keep at their core the notion of an approach to impact assessment that is shaped by humanitarian principles first and foremost. At the heart of this ongoing effort should be a principled and pragmatic approach to realising the potential of impact assessment to contribute to improving relief efforts, and delivering better assistance to those affected by emergencies around the world.

2.1 Introduction: background, aim and methodology

Before the 1990s, few humanitarian organisations thought to measure the consequences of their actions, assuming that just the provision of their assistance was evidence of its benefit (Barnett, 2005; Crisp, 2004). As Barbara Harrell-Bond observed in 1986:

‘Humanitarian work... is thought to be selfless, motivated by compassion, and by its very definition suggests good work... As relief is a gift, it is not expected that anyone (most especially the recipients) should examine the quality or quantity of what is given’ (Harrell-Bond, 1986).

This attitude has changed dramatically over the last two decades. In a context of limited resources and growing interest in and concern about humanitarian performance, there is now much more analysis and scrutiny of humanitarian operations. This is demonstrated by the growing number of evaluations and accountability mechanisms applied within the sector (Clarke and Ramalingam, 2008). As Crisp (2004) notes, the evaluation of humanitarian action has become ‘big business’, attracting new levels of donor funding and agency commitment, as well as political and public interest.

However, despite these developments, evidence about the impact of aid interventions (humanitarian or development) is not forthcoming, and impact assessment, nominally the day-to-day work of evaluation departments, is held by some critics to be insufficiently rigorous in the aid sector as a whole (CGD, 2006; Forss and Bandstein, 2008). Consequently, interest in humanitarian impact assessment is mounting, characterised by a shift in the focus of evaluations from delivery processes and how funds are spent, to analysis of the effects of humanitarian assistance on the lives of the people it is intended to serve.

This new interest is propelled by a number of factors including a broad shift to a language and practice of managing and being accountable for results, calls for more ‘evidence-based’ policy, methodological advances for assessing causality, increasing awareness of the potential unintended, negative effects of humanitarian aid and an increasingly urgent need to generate knowledge about ‘what works’ (NONIE

Subgroup 2, 2008). Aid agencies and convening bodies are now committing to assessing humanitarian impact in both explicit and implicit ways.

For example, several initiatives have been set up to monitor and evaluate humanitarian outcomes and impacts, including Collaborative for Development Action's Listening Project; the Fritz Institute's Beneficiary Perception Surveys; the Health and Nutrition Tracking Service (HNTRS); Quality Compass, developed by Groupe URD; the Standardized Monitoring and Assessment of Relief and Transitions (SMART); and the Tsunami Recovery Impact Assessment and Monitoring System (TRIAMS). Moreover, many relief agencies, including WFP, FAO, ECHO, UNICEF, UNHCR, OXFAM GB, CARE International and the IRC, now include impact in their evaluation guidelines (Watson, 2008; WRC, 2008). The Emergency Capacity Building project, a coalition of INGOs, has led the development of the Good Enough Guide (ECB, 2007). In the wider aid sector, important work is being led by the OECD-DAC Evaluation Network, the Network of Networks on Impact Evaluation (NONIE), and the International Initiative on Impact Evaluation (3IE).

Despite some progress in this area, however, humanitarian impact assessments have not become common practice. Of the respondents to an ALNAP survey of full and observer members in early 2009, the majority had conducted between one and five impact assessments in the past four years. While most respondents were confident that impact assessment was possible, a common reason given for its low frequency was that humanitarian impact was complex and difficult to assess. Moreover, many respondents pointed to a significant degree of confusion regarding the definition and purpose of impact assessments: 'most organisations that claim to do impact assessments do not really appreciate what the terminology means – rather they feel that any measurement of results is a measurement of impact' (ALNAP, 2009).

The humanitarian sector has yet to establish a common understanding of 'impact' and how to measure it. Impact-assessment approaches, methods, tools, resources and management structures differ widely between organisations. In 1999, Niels Dabelstein wrote that, 'the widely varying quality, scope and depth of these evaluations may be likened to a "methodological anarchy"' (OECD-DAC, 1999, pp. 2). This remains true of impact assessment today. The lack of shared understanding has muddied the discussion about the desirability and feasibility of impact assessments in different contexts, and muddled the selection of appropriate methodologies. Moreover, the role of beneficiaries highlighted by Harrell-Bond twenty years ago remains at best partial in the evaluation of humanitarian aid (Kaiser, 2004).

The purpose of this chapter is to provide an overview of current debates, issues and challenges concerning humanitarian impact assessment. Its overall goal is to clarify issues relating to the application and use of impact assessment as a tool for promoting accountability and learning in the sector, and to contribute to building consensus among the humanitarian community about directions for improving humanitarian impact assessment and, ultimately, performance. The chapter is the final product of a programme of research conducted between September 2008 and April 2009, including discussions and debates at the 24th ALNAP biannual meeting in Berlin in December 2008¹.

A literature review and key informant interviews carried out in autumn 2008 identified a number of challenges commonly associated with humanitarian impact assessment. These were presented at the ALNAP 24th biannual, which gathered actors from the humanitarian and development sectors to discuss and develop a shared understanding of the limits and possibilities of humanitarian impact assessment. Following the meeting, and in consultation with an advisory group, case studies were identified, partly through soliciting ALNAP members to share examples of impact assessments, and partly through further interviews and web-based database and journal searches.² Further information on this process is included below in Section 2.3.1, Overview of the case studies.

Lastly, a survey was sent to full and observer members asking them about the potential benefits and challenges of conducting impact assessments within their organisations. Although the response rate was very good, with almost ninety respondents, this should not be seen as representative of the sector as a whole, but as providing additional, complementary information to the findings from the literature review, case studies and key informant interviews.

The next section (IA.2) outlines a framework for thinking about the various conceptual, methodological, political and institutional challenges of assessing humanitarian impact. This framework is then used to examine four case studies of humanitarian impact assessment in practice (Section IA.3). This examination seeks specifically to draw lessons from an exploration of the practical ways in which different humanitarian impact-assessment initiatives have attempted to overcome the challenges outlined in Section IA.2. The final section (IA.4) presents conclusions and offers recommendations for how future work in this area could improve on existing practice.

2.2 A conceptual framework for humanitarian impact assessment

The conceptual framework for thinking about humanitarian impact assessment presented in this section of the chapter includes a number of inter-related areas, which should all be considered when thinking about, designing or implementing impact assessments. These include:

- 1 Understanding and balancing stakeholder interests: who wants humanitarian impact assessment and why?
- 2 Understanding and defining humanitarian impact
- 3 Methodological approaches and challenges: indicators, attribution, baselines, monitoring, time and timing
- 4 The importance of engaging local actors and affected populations
- 5 Capacities and incentives for improved humanitarian impact assessment

2.2.1 Understanding and balancing stakeholder interests

There is a wealth of evidence showing that evaluations – including impact assessments – are more likely to be used if their design and process fits the purpose and interests of the end users (Patton, 1997; Sandison, 2006; Leeuw and Vaessen, 2009). It is therefore always essential to ask, ‘who wants to know about humanitarian impact and why?’ Questions of whether or not humanitarian aid works, and what can be done to improve it, are of interest to a wide range of stakeholders including UN agencies, donor governments and publics, recipient

governments, international and national NGOs, the media and those affected by disasters.

This diversity means there is no single consistent view of impact assessment – different stakeholders have very different views about what impact assessments are for. As Howard White (2007) observes, the term has taken on a range of different meanings in the overall aid sector, including:

- beneficiary assessment, identified as beneficiary or participatory impact assessment
- studies focusing on final welfare outcomes
- evaluations carried out some time after an intervention has finished
- studies encompassing a whole country or sector
- counterfactual analysis – assessing results with and without an intervention
- studies with a specific focus, such as ‘environmental impact analysis’.

It is unsurprising, therefore, that humanitarian impact assessments vary widely in their ambition and scope. Some may attempt to evaluate the results of a single project with clear objectives. Others may aim to measure the overall impact of a country-wide programme, an organisation, or the humanitarian sector as a whole cutting across sectors, themes and geographic areas (Chapman and Mancini, 2008; Hofmann et al., 2004; NONIE Subgroup 2, 2008). To date, however, the humanitarian sector has tended to focus on project-level impact assessment. This is partly because this is arguably easier than evaluating the impact of whole programmes or complex systems (Stein, 2008), but may also be because of a general tendency to focus performance-related activities on the “lower end” of the system (Ramalingam et al., 2009)

Like all humanitarian evaluations, impact assessments predominantly relate to two institutional priorities: accountability and learning. Yet, there is an inherent tension between these two objectives. For example, the framing of impact assessment in terms of accountability for results can encourage risk aversion among humanitarian

Table 2.1 Who wants to know about humanitarian impact and why?

Scope of impact assessment	Who wants to know and why?
Impact of projects	<p>Aid agencies in order to improve their work, demonstrate impact and make choices between different strategies and scale-up programmes that work</p> <p>Donors to choose what to fund and to develop policy</p> <p>Recipient governments to guide disaster preparedness, planning and response</p> <p>Affected populations to voice concerns about needs and ensure that the knowledge generated is accurate and used in ways that improve their situation</p>
Impact of organisations	<p>Aid agencies to demonstrate success and raise money from the public and from donors</p> <p>Governance structures may require information about impacts in order to assess the strategic direction of the organisation</p> <p>Donors to choose between competing agencies or to make choices about whether to use NGOs or private contractors</p> <p>Recipient governments to choose who to register and work with as partners</p>
Impact at a sectoral level	<p>Aid agencies and donors to build up the evidence-base for what works; to develop sectoral policies and best-practice protocols and guidelines</p> <p>Recipient governments to develop and implement sectoral policies and best-practice protocols</p>
Impact at country level or for a particular crisis	<p>Donors to know how many lives were saved, livelihoods supported, etc</p> <p>Recipient governments to learn about which policies work and which do not; to assess whether appealing for international aid was the right thing to do</p> <p>Agencies to learn about and improve programming; to advocate for increases in levels of aid</p>
Impact of international engagement in a crisis, including, but not limited to, humanitarian aid	<p>Donor governments to review their overall engagement with countries in crisis (diplomatic, political, military and aid)</p> <p>Aid agencies to advocate for greater political engagement in ‘forgotten crises’</p> <p>Governments to promote the coherence of political and humanitarian agendas</p> <p>Aid agencies to maintain the neutrality and independence of assistance</p>
All levels	<p>Affected populations to ensure that the knowledge generated by impact assessments is accurate and used in ways that improve their lives, and to hold governments and agencies to account</p> <p>Donor publics to know whether the money they donated made a difference, and hold to account those in whose name funding is raised</p> <p>Campaigners and think-tanks for advocacy influencing the formulation of policy and best-practice guidelines for humanitarian programming</p> <p>Academics in order to further research interests</p> <p>The media to hold to account those in whose name funding is raised, or to sell stories</p>

Source adapted from Hofmann et al., 2004

agencies and thereby undermine the learning and innovation required for improving performance (Stein, 2008; OIOS, 2008). The two distinct purposes underlying impact assessment have been described as ‘proving impact’ (accountability) and ‘improving practice’ (learning).

It may be that the interests of diverse stakeholders are not all reconcilable and achievable within a particular impact assessment. As Peta Sandison notes, part of the problem facing humanitarian evaluations (including impact assessments) is that too much is expected of them by too many different stakeholders. In her chapter on evaluation utilisation in a previous *Review of Humanitarian Action*, she observes, ‘if we continue to expect evaluation to cover most of the accountability [and learning] needs of the sector, we will be disappointed’ (Sandison, 2006).

Choices regarding the purpose and scope of impact assessments are political and have important implications for the selection of appropriate methodologies, the kinds of knowledge and conclusions generated, and follow-up and use of these. It is crucial therefore that adequate time is made for the meaningful participation of all stakeholders in defining the purpose and scope of impact assessments (Patton, 1997; Sandison, 2006). Table 2.1 presents a range of stakeholders and the reasons why they might be interested in impact assessment. It is intended to be illustrative, rather than exhaustive, and provides a starting point for clarifying needs and interests around impact assessment.

Key points on understanding and balancing stakeholder interests

- 1** Findings from impact assessments are more likely to be used if the design and process fits the needs and interests of the end users. It is therefore always essential to ask, ‘who wants to know about humanitarian impact and why?’
- 2** Impact assessments may be of interest to a wide range of stakeholders including UN agencies, donor governments and publics, recipient governments, international and national NGOs, the media and those affected by disasters.
- 3** Choices regarding the purpose and scope of impact assessments will shape the selection of appropriate methodologies, the kinds of knowledge and conclusions generated, and follow-up and use of these. It is therefore essential that adequate time is made for the meaningful participation of all relevant stakeholders in defining the purpose and scope of impact assessments.

2.2.2 Understanding and defining humanitarian impact

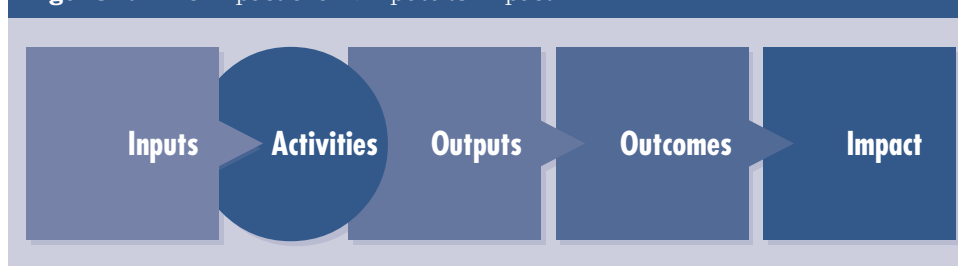
What is humanitarian action?

There is an ongoing and active debate among humanitarian agencies about what constitutes ‘humanitarian action’. Understandings and definitions are highly dependent on how individual agencies view the work that they do, whether their focus is on short term relief or extends to a broader range of interventions including protection, livelihood support, peace-building, advocacy and human rights. This lack of consensus does not necessarily present a serious challenge to impact assessment, as what matters most is how agencies understand and define impact within the context of the work that they do.

What is humanitarian impact?

The widespread adoption by donor agencies of Logical Framework Analysis during the 1990s and more recently of Results Based Management (RBM) has encouraged both humanitarian and development evaluation communities to view ‘impact’ as the final step in a linear ‘chain’ of results. This conceptual approach suggests that an intervention’s inputs lead to activities, which generate outputs, and then outcomes (often called effects or results), and which ultimately lead to impacts (Figure 2.1). Each of these terms have been defined by the Organisation for Economic Cooperation and Development/Development Assistance Committee (OECD-DAC) as shown in Box 2.1

Figure 2.1 The impact chain: inputs to impact



The causal pathway this linear model depicts is undoubtedly over-simplistic and as such has been subject to a number of critiques (Hofmann et al., 2004; Pawson, 2003; Watson, 2008). In reality, the causal links between humanitarian interventions and

Box 2.1 The OECD-DAC definition of impact

The OECD-DAC's *Glossary of Key Terms in Evaluation and Results Based Management* defines the components of the results chain as follows:

- **Inputs** The financial, human and material resources used for the development intervention.
- **Activity** Actions taken or work performed through which inputs, such as funds, technical assistance and other types of resources are mobilised to produce specific outputs.
- **Outputs** The products, capital goods and services which result from a development intervention.
- **Outcomes** The likely or achieved short-term and medium-term effects of an intervention's outputs.
- **Impacts** Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.³

Source OECD-DAC, 2002 (http://www.oecd.org/findDocument/0,2350,en_2649_34435_1_119678_1_1_1,00.html)

their outcomes and impacts are more messy, unpredictable and iterative. Nevertheless, used sensibly, this model is a *useful starting point* for mapping the causal assumptions underlying the specific objectives of an intervention, by providing a common framework to think about how the intervention will actually deliver the changes we wish to bring about.

However, this model is not always used in particularly thoughtful ways. Impact assessment is complicated by the fact that the underlying logic or 'theory of change' of humanitarian interventions is often overly simplistic (i.e. linear) or is poorly thought through, unclear or overambitious.

In many cases, interventions may not be based on sound evidence of needs (Watson, 2008; NORAD, 2008; Darcy and Hofmann, 2003). This can make the later process of

measuring and judging impacts quite challenging. Addressing participants at the ALNAP biannual meeting in December 2008, James Darcy, Director, Humanitarian Policy Group, ODI, emphasised the importance of linking needs assessments to impact assessments, describing the two as ‘inextricable’. He noted:

‘It is easy to misrepresent a problem. In southern Africa in 2003, WFP claimed that 14 million people were at risk of starvation. The donors accordingly funded a big regional food aid programme, and 14 million people did not die. What should we conclude? Either the programme was highly successful or else the original prognosis was wrong. It was indeed demonstrably wrong. This happens all the time.’

Another key challenge facing humanitarian impact assessment is the lack of clarity surrounding the humanitarian definition of ‘impact’. This needs to be considered both in general terms, in the overall context of humanitarian work, as well as in more specific terms, when defining impact precisely in relation to a particular humanitarian intervention. While there is no generally accepted definition of humanitarian impact, one of the most commonly used is:

‘Lasting or significant changes – positive or negative, intended or not – in people’s lives brought about by a given action or series of actions’ (Roche, 2000).

Unpacking this definition is helpful in informing key debates around humanitarian impact assessment. There are three main points worth highlighting.

First, this definition was initially adopted in the late 1990s by Oxfam GB in an attempt to overcome the perceived shortcomings of the definition of the OECD-DAC (as highlighted in Box 2.1 earlier). This emphasises ‘long-term’ effects and is therefore perceived as inadequate for capturing the shorter-term and immediate nature of many of the hoped-for impacts of humanitarian work (Roche, 2000).⁴

In identifying impact as ‘lasting or significant change’, Roche’s definition still conforms to the notion that impact is the final result of an intervention, but breaks away from the notion that this is necessarily long term. This should not be seen as conflating outcomes and impacts but, rather as drawing attention to the fact that the timeframe within which impacts emerge may be shorter in humanitarian contexts than in development contexts. As will be seen later in the case study section, the

actual timeframe in which impacts become visible depends partly on the goals and objectives of an intervention and partly on the wider context.

Second, Roche's definition is useful because it brings together three key elements: *changes*, which take place in *people's – aid recipients – lives* and *attribution* of these changes to specific activities or actions. Although these three elements may seem obvious, together they help to distinguish the focus of impact assessment from that of traditional aid evaluations, which have tended to focus on outputs, such as numbers of schools built or blankets delivered. Such limited approaches have proved inadequate for capturing the real changes in people's lives that have occurred as a consequence of receiving humanitarian aid.

It is worth pointing out that in practice, some humanitarian interventions might not focus on 'changes in people's lives' but on differences relative to what would have happened without the programme. For example, in some situations, food and cash distributions which lead to overall improvements in household food security would be viewed as having a positive impact. However, in other situations, distribution may focus on the *preservation* of existing levels of food security. Offsetting an anticipated fall in food security could equally be seen as a positive impact, as long as there is convincing evidence that the situation would have deteriorated without the intervention.

The third important element in Roche's definition is the broadening of impact beyond simply considering whether objectives have been achieved or measuring direct effects on intended beneficiaries. The linear model shown in Figure 2.4 can lead to 'results'- or 'objectives'-based approaches to analysing impact, focusing solely on a set of pre-defined institutional objectives (Hospes, 2008; Bolton et al., 2007). By defining impact as potentially *positive, negative, intended or unintended*, Roche recognises that processes of social change are rarely solely the product of a managed process undertaken by humanitarian agencies. Rather, they are the result of wider, complex historical, social, political, economic and environmental processes, and can include unintended effects on families, households and communities.

Taking these wider contextual factors into account is considered particularly important in emergency contexts to ensure that the principle of 'do no harm' is maintained (Watson, 2008). This is sometimes articulated as the difference between assessing '*effectiveness*' which focuses on whether the intermediate objectives of the intervention have been achieved, and '*impact*' which is perceived as going beyond


intervention-planning documents to consider the part an intervention plays in wider socio-economic and political contexts (Beck, 2006).

There is, and will continue to be, debate about the value of generic definitions of impact such as the OECD-DAC or Roche definitions provided above. There is general acknowledgement of the importance of going beyond ‘effectiveness’ to examine the wider, unintended impacts of humanitarian operations. However, the general view among key informants interviewed for this study is that a rigid definition of humanitarian impact as ‘long-term effects’ or ‘lasting change’ may be unhelpful because: a) some humanitarian impacts may reveal themselves within short timeframes; and b) not all humanitarian goals are intended to be sustainable.

Of the survey respondents (ALNAP, 2009), over half defined impact in a similar way to Roche, broadly describing it as the overall ‘change’ resulting from a project, often incorporating outcomes. In the words of one respondent, ‘my understanding is that impact is the changes (positive and negative) that have happened as a result of the intervention’. Around a third of respondents used terms similar to OECD-DAC as above, defining impact as ‘long-term’ effects, and distinguishing impacts from outcomes by reference to timescale. Explicit reference to the importance of attribution was made in a minority of responses.

Clearly, generic definitions can be useful, but only up to a point. More important is the ability to use such generic definitions in a thoughtful, purposeful way to help think through the hoped-for changes to be brought about by a given intervention. Any *practical and operationally useful* definition of impact – how it results from given actions, and how it differs from outcomes – will by necessity depend on the specific context and the goals of the humanitarian programme being assessed.

Key points on defining humanitarian impact

- 1 If the underlying logic or ‘theory of change’ of a humanitarian intervention is unclear, overambitious or not based on need, this makes it difficult to measure and assess the intended results.
- 2 While there is no generally accepted definition of humanitarian impact, Roche’s definition (stated above) is useful because it: 

- ▶ • captures the shorter-term and immediate nature of many humanitarian impacts
 - encompasses three key elements – changes, people’s lives, and attribution of those changes to specific actions
 - focuses on both intended and unintended effects.
- 3** Generic definitions are useful, but most important is the ability to use such definitions in a reflective, purposeful way to think through the hoped-for changes to be brought about by a given intervention. Any practical and operationally useful definition of impact will depend on the specific context and goals of the programme being assessed.

2.2.3 Methodological approaches and challenges

Indicators: moving beyond outputs

Once impact is defined, the identification and measurement of relevant impact indicators is a crucial part of gathering evidence of the impact of an intervention. However, the appropriate forms and means of impact measurement in relation to humanitarian action are highly contested. Impact assessment inevitably involves value judgements about which kinds of changes are significant for whom (Roche, 2000). The different motivations of different stakeholders for impact assessment (Section 2.2.1) can result in distinctive criteria for assessing positive or negative impacts. It is therefore important to involve beneficiaries in identifying appropriate and objective indicators for impact, as these may differ from those of other stakeholders (also see Section 2.2.4).

Although the terminology sometimes varies, there are generally two types of indicator: those relating to the implementation of programmes (input, process and output indicators) and those concerned with programme effects (outcome and impact indicators). These are shown in Table 2.2. Humanitarian agencies tend to use a mix of indicators depending on their own monitoring and reporting systems, and on the particular function of the indicators collected (Hofmann et al., 2004).

Table 2.2 Types of indicator: the example of measles-immunisation programmes

Implementation of the programme			Effect of the programme	
Input indicator	Process indicator	Output indicator	Outcome indicator	Impact indicator
No. of vaccines administered	No. of people trained	Percentage vaccinated	Measles cases decrease	Mortality decreases

Source Hofmann et al, 2004

Historically, humanitarian agencies have tended to collect *process* rather than *impact* indicators. For example, the Sphere indicators – probably the most comprehensive attempt to define standards and indicators for five key sectors of humanitarian aid (Sphere, 2004) – focus largely on minimum output levels, and are not designed to address impact. Monitoring and evaluation of processes, products and services is an important management function and is also essential for understanding how outcomes and impacts come about (Catley et al., 2008) – Sphere has tremendous potential value in this regard.

However, as Elliot Stern, Professor of Evaluation Research, Lancaster University, pointed out in his keynote speech at the ALNAP biannual meeting on Impact Assessment, ‘the balance of evaluative effort is often skewed towards processes that are not linked to outcomes or impacts’. While almost three-quarters of survey respondents stated that they measured impact-related indicators, a similar percentage (67 per cent) found this ‘very challenging’ or ‘challenging’. The importance of moving beyond ‘outputs’ to ‘outcomes’ is usefully and humorously illustrated in the quotation from Quinn-Patton in Box 2.2.⁵ However, in cases of strong evidence that an output leads directly to an outcome or impact, for example with polio vaccinations, process indicators may be sufficient.

It is also important to distinguish ‘proxy’ impact indicators from ‘real’ impact indicators. As researchers at the Feinstein Institute note:

‘If a project provides training in new and improved farming practices, a transfer of skills and knowledge or human capital would be expected. Only when applied might this knowledge transfer translate into improved crop yields, resulting improved household income from increased crop sales and improved household food security. The application of knowledge and the

Box 2.2

Moving from outputs to outcomes

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

Source Quinn-Patton, 1997, pp. 157–158

improved yields attributable to this project are therefore only *proxy indicators of impact*. If some of the extra food produced is consumed by the farmer and his family, this use represents a ‘real’ food security and nutritional benefit, or livelihoods impact. Alternatively, if increased income derived from crop sales allows for livelihoods investments in health, education, food and food production, or income generation, these expenditures would also represent ‘real’ impacts on the lives of project participants’ (Catley et al., 2008 p. 20).

The researchers add that ‘impact indicators look at the end result of project activities on people’s lives. Ideally, they measure the fundamental assets, resources and feelings of people affected by the project. Therefore, impact indicators can include household measures of income and expenditure, food consumption, health, security, confidence and hope’, (Catley et al., 2008 p. 20). Viewed in this way, proxy indicators

of impact may well provide a sound basis for understanding outcomes as an intermediate stage between outputs and impacts depending, of course, on the definition of ‘impact’ and theory of change adopted by the project.

Key points on indicators

- 1** Impact assessment inevitably involves value judgements about which kinds of changes are significant for whom. It is therefore important to involve aid recipients in identifying appropriate and objective indicators for impact, as these may differ from those of other stakeholders.
- 2** It is important to move beyond simply measuring process indicators, and towards measuring outcome and impact indicators. However, using a mixture is ideal as process indicators help to explain how outcomes and impacts are realised.
- 3** Proxy indicators of impact may tell you whether the objectives of a project were accomplished, but ‘real’ impact indicators focus on the changes in people’s lives resulting from those objectives being met.

Overcoming the attribution problem with appropriate approaches and methods

A central element of impact assessment is the analysis and determination of causality. This is the process by which impact evaluators identify a causal relationship between a particular intervention or activity and a particular change in the conditions of people’s lives. In order to do this, evaluators need to be confident that the change observed is caused by the intervention and not by other co-existing factors, which can ‘confound’ the analysis (Fearon et al., 2008). This is commonly referred to as the *attribution problem*, and is the focus of an extensive body of literature in the evaluation field.

The disrupted, stressed and fluid characteristics of the post-disaster and conflict-affected contexts in which humanitarian agencies operate are especially resistant to attempts to establish causality and attribution. Most operations involve dozens of agencies, usually involving some combination of: the national and local authorities of

the affected country, UN agencies, international NGOs, national NGOs and in some contexts, national and foreign militaries. In addition to these ‘external’ efforts the affected communities themselves are invariably mobilising resources and providing their own networks of support and care (Borton, 2008). Faced with such levels of complexity, many humanitarian agencies struggle to attribute outcomes and impacts to their interventions. There are however a number of ways in which agencies can overcome these difficulties.

Comparative and theory-based approaches

Methodological approaches used to establish causality and overcome the attribution problem can be grouped into two broad types.

1 Comparative approaches

Comparative approaches include a range of quantitative methods including quasi- and full (randomised) experiments (see Box 2.3) and attempt to establish a counterfactual – that is what would have happened if an intervention had not taken place. Counterfactuals may be either geographical (comparing a site or group in receipt of aid with one that is not) or temporal (comparing circumstances ‘before and after’ an intervention takes place).

Given the complexity and fluidity of humanitarian contexts, establishing a geographical counterfactual using a control or comparison group – a group similar to the group benefitting from an intervention and otherwise subject to the same influences (Roche, 1999) – can be extremely challenging. Numerous factors might explain an observed difference in outcomes or impacts between communities that do and do not receive a particular humanitarian intervention. For example, if comparison groups start off richer or poorer, or more or less affected by a disaster, then estimates of a programme’s impact could mistakenly reflect these pre-existing features (Fearon et al., 2008). Moreover, these pre-existing features may have led to the community or group to be selected into the experiment in the first place; a bias known as ‘selection bias’. One method currently being strongly promoted by some as a ‘gold standard’ in impact assessment, given its potential for overcoming these problems, is that of randomised evaluation (See Box 2.3 on pages 26–29 for further information on this).

Box 2.3 Are randomised controlled trials (RCTs) superior to other methods?

In an RCT, individuals, families or communities are randomly assigned to one or more interventions or ‘treatments’, and to a control group. By measuring the treatment and control groups before and after a program, researchers claim they can isolate the effects of the program from other factors.

RCT protagonists argue that by eliminating all spurious explanations for results, RCTs have a unique advantage over ‘quasi’-experiments and other theory-based impact assessment methods for assessing impact. This perceived unique advantage stems from two distinguishing features of RCTs:

- a) random assignment and
- b) a large sample size.

Respectively, these features eliminate selection bias and ensure that other confounding factors (both observable and unobservable) are “balanced out” between ‘treatment’ and ‘control’ groups.

While the randomisation process can be seen as a special virtue of RCTs in that it does eliminate selection bias, the following list provides a number of reasons why RCTs may be no better than quasi-experiments or other theory-based methods when applied to humanitarian contexts:

- The RCT designs carried out in aid contexts are not ‘fully’-blind, which means that evaluators and subjects know who the ‘treatment’ and ‘control’ groups are. As such, evaluators may be biased when assessing results and/or subjects may change an aspect of their behaviour in a way that affects outcomes or impacts, simply because they know are being studied.⁴⁰
- It is impossible to tell how much the specific context of an RCT shapes the outcomes and impacts observed. The results of RCTs must therefore always be backed up by high quality contextual information from the field, gathered through non-RCT methods.

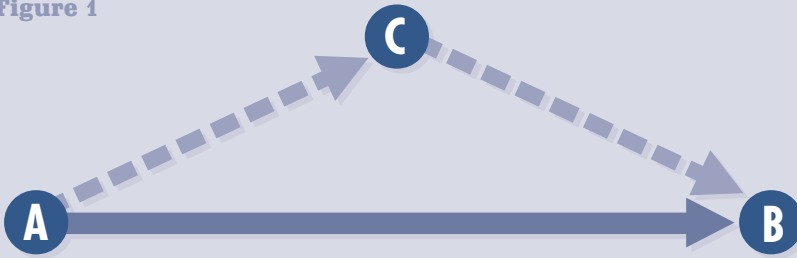
CONTINUED

Box 2.3

Are randomised controlled trials (RCTs) superior to other methods? *continued*

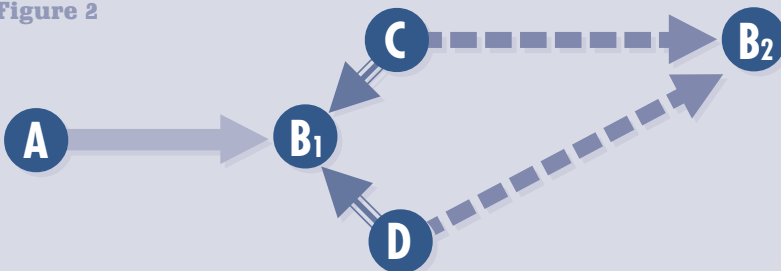
- Establishing causality is more complex than establishing a counterfactual. In complex societies, there may be causes which are not immediately visible within the linear logic model of an RCT, as illustrated below.

Figure 1



In this situation, an RCT may reveal that intervention A causes effect B. However, intervention A only causes B provided that there is some interaction with circumstance C, which is “lurking in the background”. However, circumstance C may not be visible within the logic model characteristic of RCTs.

Figure 2



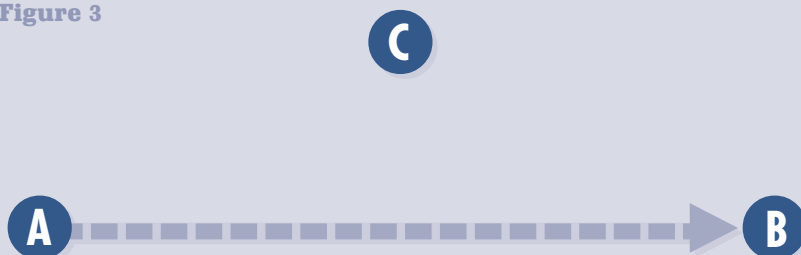
In this situation, event A appears to cause effect B. However, in fact, intervention A has nothing to do with effect B, which is caused by a number of contextual and other factors that are quite separate from event A. Indeed, there is another B(2) going on somewhere else which is also not linked to intervention A.

CONTINUED

Box 2.3

Are randomised controlled trials (RCTs) superior to other methods? *continued*

Figure 3



In this situation, not only does intervention A not lead to effect B, but circumstance C also has no effect. In a similar way to Fig. 4, this example highlights the possibility that event A might only lead to effect B in the presence of certain mediating factors (circumstance C or others), which themselves only come into force under specific circumstances.

The above points demonstrate that many of the unique credentials of RCTs do not hold when they are applied in ‘real world’ contexts. Moreover, the large numbers of subjects required by RCTs for ensuring that confounding factors are evenly distributed between ‘treatment’ and ‘control’ groups can make them costly compared with quasi-experiments, which require fewer subjects (Scriven, 2008).

Many in the humanitarian community perceive RCTs to be unethical as they contradict humanitarian principles about serving those in need. In response, RCT proponents argue that whenever funds are limited, or programmes need to be expanded in phases, only a portion of potential beneficiaries can be reached at any time. Choosing who initially participates by lottery is therefore no less ethical (and perhaps more so) than many other approaches (CGD, 2006; Nelson, 2008).

As Scriven observes, the bottom line is that, “...when we total up the strengths and weaknesses of any of the half-dozen leading [impact assessment] designs

CONTINUED

Box 2.3

Are randomised controlled trials (RCTs) superior to other methods? *continued*

we get the same situation, namely each has substantial entries in both columns: there is no edge for [real world] RCTs that holds across all cases" (2008, pp. 48).

Even advocates of RCTs admit that they are not suitable in all contexts. According to Esther Duflo of MIT Poverty Lab (cited in Stern, 2008), 'randomised evaluations... are only suitable for programmes that are targeted to individuals or communities, and where the objectives are well defined'. The European Evaluation Society (EES, 2008) is more specific, stating that RCTs are suitable only in contexts where:

- a linear causal relationship can be established between the intervention and desired outcome
- it is possible to 'control' for context and other factors
- it can be anticipated that programmes under both experimental and control conditions will remain static for a considerable time
- randomness and chance has no influence
- it is ethically appropriate to engage in randomisation.

Some of those interviewed for this study suggested that RCTs may be most relevant in contexts where there is a difference of opinion in the aid community about the merits of one type of intervention over another, for example 'does cash or food work better'? The question could be answered by randomly assigning cash and food to different groups and comparing the results. Another context in which RCTs could be useful is where a variation of a programme design is randomly assigned within a programme. This potentially overcomes the ethical issue of depriving people who are vulnerable or in need. This demands further research.

Source Stern (2008); Scriven (2008).

By contrast, a weakness of temporal ‘before and after’ comparisons is that the circumstances before an intervention may not provide an adequate proxy for how the situation would have evolved in the absence of an intervention. However, there are some cases in which there are no other plausible explanations of changes, so before versus after will suffice. A good example is the impact of a water supply project on the time household members spend collecting water. The average time falls after the project. The only plausible explanation is the improved proximity of water. In this case there is no need for a comparison group from outside the project area. The counterfactual is the time spent collecting water before the intervention (White, 2009).

2 Theory-based approaches

Theory-based approaches including theories of change, case studies, causal modelling, outcome mapping, most significant change and ‘realist’ evaluation (Jones et al., 2009), involve looking at a particular case in depth and theorising about the underlying causal links between actions, outcomes and impacts.⁶ The idea is to build a programme theory of change.

In practice, in complex settings, ‘often the most that can be done is to demonstrate through reasoned argument that a given input leads logically towards a given change, even if this cannot be proved statistically’ (Roche, 1999, p. 33).

One of the most straightforward approaches is General Elimination Methodology, which involves three steps:

Step 1 Identifying the list of possible causes for the outcomes and impacts of interest.

Step 2 Identifying the conditions necessary for each possible cause in the list to have an effect on outcomes or impacts.

Step 3 Working out whether the conditions for each possible cause are present or not.

Working through these elimination steps means that the final set of causes include only those causes whose necessary conditions are completely present (Scriven, 2008).

Theory-based methods require that causality must be inferred from information about the context, from beneficiaries and key informants and by triangulation with data from other sources. However, evaluators of any kind do not often solicit this sort of information: ‘we always say “context matters”, but what is remarkable is how little effort has been spent studying contexts’ (Stern, 2008).

It is important to note that comparative and theory-based approaches are not mutually exclusive. For example, there is no reason why a randomised evaluation could not be embedded within theory-based approach. However, at present, most impact assessments do not combine the approaches (White, 2009).

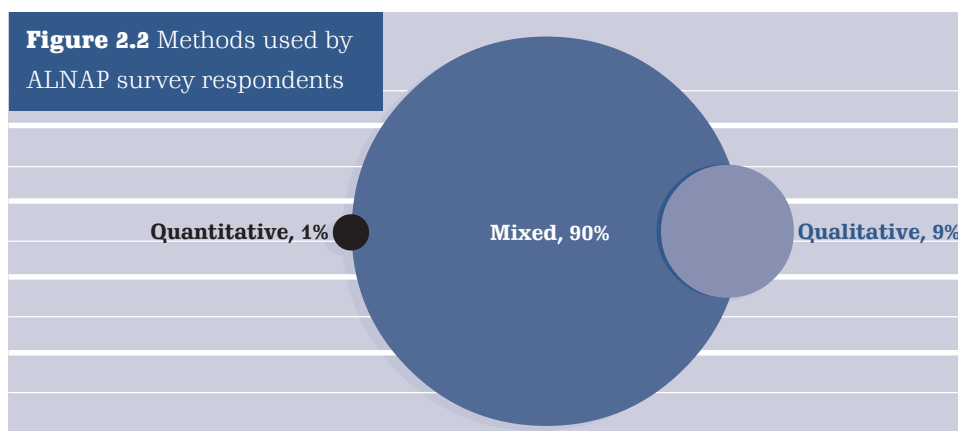
Quantitative and qualitative methods

The debate about the value of comparative versus theory-based approaches is frequently conflated with the debate about quantitative versus qualitative methods. It is important to highlight that this is an artificial conflation of two distinct issues, comparative approaches can involve either quantitative or qualitative methods or a combination, and theory-based approaches will usually involve a mixture of the two.

What is clear is that quantitative and qualitative methods are suited to answering different types of evaluation questions. Quantitative methods are better for ‘what’ and ‘where’ questions, while qualitative methods are better for ‘why’ and ‘how’ questions and are good at capturing processes (Prowse, 2007). The two are often more powerful together than alone – qualitative methods can be used to identify quantitative indicators or explain quantitative findings, while quantitative methods can be used to enhance a qualitative study or in parallel with qualitative data (Steckler et al., 1992). A strict methodologically determined approach to impact assessment can often hide more than it reveals.

The results of the ALNAP survey reveal that the great majority of respondents use a mixture of qualitative and quantitative methods when carrying out impact assessments (Figure 2.2). Asked to explain their answers, 59 respondents (of 77 who responded to the original question) provided additional comments. By far the most common reason given for using mixed methods (81 per cent) was the need for triangulation in order to achieve a representative picture.

Figure 2.2 Methods used by ALNAP survey respondents



Selecting appropriate approaches and methods

The key to doing good impact assessments is to ensure that selected approaches and methods fit the specific circumstances of an impact assessment – its definition, purpose, questions, context, scope, indicators, level of existing knowledge and available resources – cannot be overemphasised. However, this is often easier said than done. Impact assessments, like any research, are shaped by both opportunities and constraints. They often involve finding an appropriate balance between the desire to understand and measure the full range of effects in the most rigorous manner possible, and the practical need to delimit and prioritise on the basis of stakeholders interests and resource constraints (Leeuw and Vaessen, 2009).

As illustrated above and later on, common trade-offs include choices between comparative and theory-based approaches, between quantitative and qualitative methods; between timing and data quality, between academic rigour and human/financial resources. All of these issues need to be carefully thought through and negotiated across key stakeholders in a particular impact assessment effort. Rather than asking, *‘which methods and approaches are appropriate for assessing humanitarian impact?’* it is more important and useful to ask, *‘which methods and approaches are appropriate, feasible and worthwhile under the specific conditions and given our goals?’*

Participants at the December 2008 biannual meeting were, on the whole, sceptical about the notion that there is any one approach or method superior to others. As noted by Nick York, DFID Head of Evaluation, ‘If impact assessment is used as a tool to reinforce [particular] methods, it is useless’. Despite this, some also voiced the concern that the current popularity of comparative approaches and experimental methods could lead to ‘methods-led’ understandings of impact, which in turn could

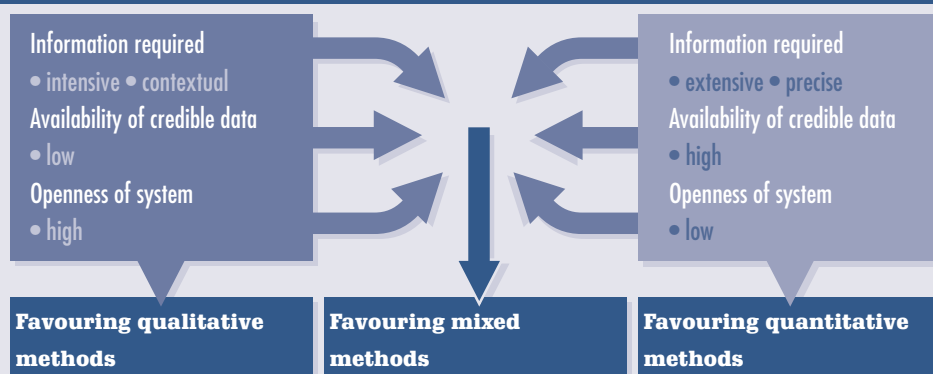
encourage the reformulation of humanitarian policies and programmes so that they become ‘evaluable’ through these methods. However, there are a number of clear reasons why such approaches should not be seen as the ‘gold standard’ for the humanitarian aid sector (see Box 2.3 on page 26–29 for more details). Instead, as the 2008 NONIE Subgroup 2 submission concluded, ‘*methodological appropriateness could be considered the “gold standard” for impact evaluation*’ (NONIE Subgroup 2, 2008, pp. ii).

The purpose and scope of an impact assessment should be taken into account when selecting appropriate methods. For example, if an impact assessment is intended to assess the effects of a discrete intervention, with a view to answering questions such as *What difference are we making? How does strategy A compare to strategy B?*, then evaluators may consider adopting a comparative approach. However, if an impact assessment is intended to assess the overall collective impact of a complex humanitarian operation involving multiple agencies and interventions, then theory-based methods may be more appropriate given the lack of a comparison group. Even if a counterfactual could be established, it would most likely be inadequate for answering questions such as *which components of the overall humanitarian effort have worked well and which have worked less well?*

In line with this, Elliot Stern (2008) usefully identifies three distinct programme contexts, and their appropriate methodological approaches:

- 1** standardised interventions in identical settings with common beneficiaries – better suited to experiments
- 2** standardised interventions in diverse settings, possibly with diverse beneficiaries – better suited to quasi-experiments and comparisons and combinations of methods
- 3** customised interventions in diverse settings with diverse beneficiaries – better suited to case studies or narrative/qualitative approaches that build plausible theories.

A complementary decision-making framework which may be worth considering in the humanitarian sector is the contingency model. This suggests that qualitative and quantitative methods are relevant in different situations, and identifies three important criteria for decision making: information required, availability of credible data, and openness of the ‘system’ being evaluated (Figure 2.3).⁷

Figure 2.3 Contingency model for methodology selection

Source Williams, 1999⁸

Key points on the attribution problem: selecting appropriate approaches and methods

- 1** Assessing impact involves determining whether changes observed are caused by an intervention or by other factors – the ‘attribution problem’. Establishing causality in humanitarian situations is complex because results are unlikely to be due to any single factor.
- 2** Two broad approaches are used to determine attribution in humanitarian impact assessment. Comparative approaches attempt to establish what would have happened without a particular intervention, and theory-based methods examine a particular case in depth to explain how an intervention could be responsible for specific changes.⁹
- 3** Selected methods and tools must fit the specific circumstances of an impact assessment – its purpose, the nature of the intervention being assessed, questions, indicators, level of existing knowledge, and resources available.
- 4** When designing an impact assessment, it is important to ask which methods are appropriate, feasible and worthwhile, and carefully consider any trade-offs, such as between quantitative and qualitative methods, timing and data quality, rigour and resources, and internal and external validity.
- 5** A mixed quantitative and qualitative approach can often provide the best information in most circumstances.

Baselines, monitoring and data collection

The last point in the methodology section illustrated the importance of basing methodological decisions about impact assessment on a number of factors, including an understanding of the availability of data. In reality, the challenge of gathering data and the absence of baseline data are often used as part of the excuse for not assessing impact.

Even when indicators have been identified and integrated across the project cycle, timely and accurate data for these may be lacking. This may be due to unstable situations, limited access to affected populations or lack of methodological rigour. Furthermore, what is available is not always reliable. A synthesis of evaluations received by the ALNAP Evaluative Reports Database in 2002 notes:

'Reports were so consistent in their criticism of agency monitoring and evaluation practices that a standard sentence could almost be inserted into all reports along the lines of: It was not possible to assess the impact of this intervention because of the lack of adequate indicators, clear objectives, baseline data and monitoring' (ALNAP 2003a, p. 107).

A chapter on 'Strengthening Monitoring in the Humanitarian Sector' in the ALNAP Annual Review (2003b) identifies three important issues concerning data focus and quality in humanitarian contexts:

- 1** monitoring systems tend to focus on collecting data at the input and output levels
- 2** data collected for monitoring purposes are mainly quantitative
- 3** data quality is often poor and poorly analysed.

A particular problem highlighted is that, while it is relatively easy to collect certain types of quantitative data and send it 'up the line', it is far more difficult to define what qualitative data to collect regularly and to analyse such data.

Moreover, the same review argues that a sole focus on quantitative data excludes a deeper understanding of the context and hence decreases opportunities for learning. According to one interviewee in the ALNAP study of monitoring: 'By providing

masses of quantitative data in reports you could distract people from the key issue of whether funding for the reconstruction of housing was the right approach in the first place' (ALNAP, 2003b). The key message here is that ongoing collection and analysis of qualitative data can be a key means of understanding causality and social process – something that has been largely missing in evaluations of humanitarian action.

Of particular importance is the issue of baselines. As the 2004 HPG Report puts it: 'Attempts to analyse the impact of humanitarian interventions are often handicapped by a lack of baseline data... It is difficult to show that a humanitarian programme has had an impact without knowing the rate at which something was occurring before the intervention began, and after it was implemented. Likewise, when people are arriving in a new location or are returning home, it is often impossible to determine the baseline before their arrival... Programmes which keep mortality low or keep water and food provision high may be successful in meeting their objectives, but if they do not have a baseline rate or a comparison group they may not be able to quantify the impact of the intervention' (Hofmann et al., 2004).

There are several ways to obtain baselines in emergencies, including 'retrospective' baselines – which use beneficiary recall to reconstruct situations before the emergency – and 'rolling' baselines' in which new beneficiaries of a programme provide background information which is compared to that of existing beneficiaries (Watson, 2008). It is also possible to construct a comparison group based on what is known about the project context and process of participant selection (NORAD, 2008). Better use of national and local data sources, where relevant, can also strengthen data quality.

Finally, particularly in humanitarian work, more effort is needed to be able to monitor outputs and collect data on outcomes and impacts in a disaggregated manner – for example, outcomes in a particular sector, in a particular geographic area, or with respect to a particular demographic group. Despite obvious trade-offs in terms of costs, disaggregated data allow stakeholders to identify more clearly gaps in assistance or emerging needs. This is particularly important if data are to be used to inform planning and decision-making. Such approaches call for more collective and better coordinated efforts in data gathering across particular emergencies, and again establishing better linkages with national surveillance systems where these are present.

Key points on baselines and data collection

- 1 Impact can be assessed despite the absence of pre-existing baseline data. There are several methods for overcoming the challenge of obtaining baselines in emergencies, including ‘retrospective’ and ‘rolling’ baselines’.
- 2 Collection and analysis of both quantitative and qualitative data on an ongoing basis is important for understanding causality.
- 3 Collection of monitoring data in a disaggregated manner – for example, according to age, gender, ethnicity, geography, sector etc. – helps stakeholders to identify gaps in assistance or emerging needs. Collective and coordinated approaches are key to doing so in an effective manner, both across agencies and with national systems where present.

Timing and amount of time

A fundamental requirement for good impact assessment is sufficient time. As one key informant noted, ‘impact assessment will never be a quick and dirty job because it is concerned with complex contexts and complex chains of causality’. While planning for impact assessments ideally should begin at the start of an initiative, this does not always happen as humanitarian assistance is often delivered rapidly in situations of acute crisis, when there are time constraints affecting all stages of the project cycle (Watson, 2008).

The timing of an impact assessment is also critical: depending on the goals of the intervention, impacts may become visible soon after an intervention or many months or even years later. Timing can also determine whether an impact assessment measures ‘proxy’ or ‘real’ impact (as discussed in the following section on indicators). There is often a trade-off to be negotiated between analysis which takes place soon after an intervention, and analysis which takes place after some time has passed, as each can enable different kinds of learning and accountability.

The literature suggests that the timing and timeframes of impact assessments often depend more on the institutional needs of agencies than on what is required for conducting useful evaluations (Forss and Bandstein, 2008). Many of those interviewed for this study perceived a lack of donor flexibility regarding timing, which was often attributed to donor ‘pre-packaging’ and short-term funding cycles.

According to one international NGO staff member:

‘Some projects have too short a timeframe to implement and so we jump steps so that we can report back to donors. The quick reporting deadlines challenge communities’ own schedules for harvesting and other events’ (CDA Listening Project, 2008).

As the case study research (Section 2.3) shows, constraints related to time and timing can result in inadequate monitoring and data collection, and ultimately reduce the quality of the findings of impact assessments.

Key points on timing and time

- 1 Impact assessments should ideally be carried out when impacts are likely to be visible and measurable. Getting the timing right depends on the specific goals of an intervention.
- 2 Insufficient time for an impact assessment can result in inadequate monitoring and data collection, leading to reduced quality of findings.
- 3 Short-term project and funding cycles may not be conducive to conducting good-quality impact assessments.

2.2.4 The importance of engaging local actors and affected populations

Ensuring the meaningful participation of affected populations is a critical element of any impact assessment process. As Roche notes, “...[an assessment of] significant and lasting change in people’s lives must take account of *their* values, priorities and judgements; projects cannot be deemed to have been a ‘success’ or ‘failure’ if the perceptions of those who the intervention aims to benefit diverge seriously from those of the project staff or an external evaluator” (1999, p. 28).

However, participation of crisis-affected populations is not yet a standard feature of evaluations and, when it does happen, it is often informal and opportunistic. According to Sandison (2006) beneficiaries of aid are now more likely to be consulted during evaluations but it is not common practice to share the findings with them

afterwards. It is only in the last few years that impetus for a change in approach has been gathering pace.

Participants at the ALNAP biannual meeting unanimously agreed the importance of incorporating the perspectives of aid recipients into impact assessments in order to get a full and accurate picture of impact. Lori Bell, Evaluation Officer, FAO, provided a useful example of how participation of affected populations can improve the quality and interpretation of findings:

‘We looked at consumption of individuals after a change in WFP aid in Mozambique. We asked [one man] for his thoughts on the programme and he told us that results were good. If we’d left it at that, we would have gotten a very clear piece of information. But, it was later revealed that he not only got food assistance from WFP, but his 3 children had moved to South Africa and were sending back remittances. He was also on a separate government assistance programme. So if we had only stayed for 5 minutes we would have thought we had succeeded. But in this case there were many other factors involved.’

However, participants also identified a number of barriers, both ‘fundamental’ and ‘pragmatic’ to participation of aid recipients in impact assessments, including:

- an imbalance of power between donors and recipients, leading to the tendency for accountability to flow ‘upwards’, excluding beneficiaries as a stakeholder group
- the fact that participation is rarely a donor requirement
- insecurity in humanitarian contexts, which can make participation challenging both for beneficiaries and the aid agency
- communication barriers, caused by language and cultural factors
- lack of time to consult adequately in life-threatening situations, compounded by the pressure to deliver
- concerns that recipient perspectives will challenge or contradict agencies’ mandates.

One innovative development in relation to this challenge is the Listening Project – a three-year project launched in 2005 by the US-based Collaborative for Development Action together with a group of interested agencies and individuals. While not strictly an impact assessment, the project is a comprehensive and systematic exploration of the ideas and insights of people on the receiving side of international humanitarian and development assistance. The first ‘listening exercise’ took place in Indonesia in 2005. Since then eleven more exercises have been held in different countries in Africa, Asia and Latin America and an exploratory listening exercise in New Orleans. According to key informants interviewed for this study, the project is motivated by a sense that “if aid providers and donors could listen carefully to recipients’ judgments of what has been useful, what has not been, and why, then they would learn a great deal about how to improve the effectiveness of their efforts.”

‘Impact’ in this context refers to the cumulative effect of aid as perceived by beneficiaries. The initiative is less concerned with *intended* impacts than with finding out whether communities got what they wanted, regardless of the original objectives or plan. In this sense, it asks, on behalf of the aid community, not just ‘did we do what we said we would do?’ but, more important, ‘Were these the right things to do?’

The Listening Project has demonstrated the value of open-ended conversations with beneficiaries, which complement quantitative surveys. It has also shown that results-based management systems encourage reporting on projects and activities against pre-determined targets or indicators that do not measure success as aid recipients would do. The head of a Sri Lankan NGO notes:

‘NGOs say in their MOUs what the indicators are. They say if we want their money, we have to take the project as is, so only we know our sadness. Indicators such as elderly women, widows, are foreign indicators given to us by outsiders. We should be able to identify some indicators ourselves – perhaps 25 per cent. Allow us the flexibility to develop our indicators. It has to come from us, not be top down... We accept their conditions because we value the aid more than getting over the restrictions’ (CDA, 2008).

However, in order to be truly participatory evaluations should not only give a voice to affected populations but empower them to contribute to both the research design and the analysis of the findings. Ultimately, beneficiaries should be actively engaged

in the design and planning of interventions that aim to improve their situation (Kaiser, 2004; Roche, 1999). For this vision to be realised, impact assessments should include affected populations, not simply as sources of valuable information, but as key actors throughout the process, from defining impact to generating indicators and baselines, to playing an active role in judging impacts.

Interestingly, a majority of the respondents to the question of beneficiary engagement in the ALNAP survey sample stated that they do involve beneficiaries in their impact assessments to some degree – whether in design and/or implementation of the impact assessment. Of these, just over half provided more detailed information on how beneficiaries were involved. Of those that presented such information, most referred to participation during implementation of impact assessments through interviews and surveys; a smaller proportion indicated that they also involved beneficiaries in the design stage⁴⁴ through focus-group discussions, surveys, key-informant interviews and mass meetings. Those who involved beneficiaries throughout the process of impact assessment mentioned the design of tools, involvement in monitoring indicators, village-level group discussions, collecting baseline data, creating community groups for the assessment, and forms of capacity building during the process (ALNAP, 2009).

Box 2.4

Learning partnerships for participatory impact assessments (PIAs)

‘The experience of the Feinstein Center shows that where project participants are included in the impact assessment process, this can create an opportunity to develop a learning partnership involving the donor, the implementing partner, and the participating communities. The impact assessment process can create space for dialogue, and the results can provide a basis for discussions on how to improve programming and where best to allocate future resources... a systematic, well designed PIA can assist communities and NGOs to measure impact using their own indicators and their own methods. It can also overcome the weaknesses inherent in many donor and NGO monitoring and evaluation systems which emphasize the measurement of process and delivery, over results and impact.’

Source Catley et al., 2008, pp. 8–10

The Feinstein International Center has recently developed a detailed approach to carrying out project-level participatory impact assessments (PIAs) of livelihoods interventions in the humanitarian sector. The PIA approach was developed partly in response to the need to carry out impact assessments in complex emergency situations with difficult operational environments (Watson, 2008), and partly in response to the lack of international standards for measuring impact in areas other than water, health and nutrition (Catley et al., 2008).¹² PIA emphasises the active role that national and local actors can play in assessing impact, and the learning benefits that ensue from a collective approach: the idea of ‘learning partnerships’ is highlighted in Box 2.4 on the previous page.

Key points on national capacity and beneficiary participation

- 1 A systematic and well-designed impact assessment should involve representatives of affected communities, as well as appropriate national and local actors, at all stages from defining impacts and generating indicators to playing an active role in judging impacts.
- 2 Participation of national and local actors and crisis-affected populations could be facilitated through the development of ‘learning partnerships’ between donors, implementing partners, communities, national actors and other relevant stakeholders.

2.2.5 Capacities and incentives for improved impact assessment

Many have noted a lack of organisational capacity within the humanitarian sector to carry out good evaluations and impact assessments, despite a wealth of tools and methods (Hofmann et al., 2004; Sandison, 2006; Watson, 2008). Considerable evidence indicates that evaluation objectives are often not defined clearly, stakeholder analysis is limited, skills in assessment methodologies, especially experimental and quasi-experimental design (CGD, 2006; Forss and Bandstein 2008; ALNAP 2003b, 2004), and in the analysis and use of impact data are lacking (Sandison, 2006; and see Section 2.3.6 on Capacities and incentives for improved impact assessment). High staff turnover, lack of a ‘learning culture’ and inadequate resources are all perceived contributing factors (Roche, 2008; Watson, 2008).

The issue of capacity is linked to incentives within humanitarian agencies and across the sector as a whole that affect motivation for and investment in impact assessment. Despite the numerous potential benefits of impact assessments for humanitarian agencies and the wider sector – including more and better evidence about what works, more robust assessments of needs and improved accountability to affected populations – the literature suggests that there are relatively few incentives that promote their application and use (CGD, 2006; Roche, 2008; Watson, 2008).

Perhaps most importantly, there is a lack of effective demand for impact assessment in the sector illustrated by the fact that funding of humanitarian assistance does not depend on evidence of positive effects or accountability to beneficiaries. While there are signs that this is beginning to change, humanitarian assistance continues without a clear sense of needing to understand effects in order to be effective. As the Tsunami Evaluation highlighted, agencies seldom go out of business for poor performance, instead and more often for poor financial performance. The system as it is currently structured provides powerful incentives for humanitarian agencies to continue as they are and not to learn or improve.

In this context, the findings of impact assessments are a public good, the benefits of which have the potential to go well beyond the organisation that generates them. This leads to a ‘tragedy of the commons’ scenario where the incentives for any single organisation to bear the significant costs do not correspond to the full collective and social benefits. Within agencies, impact assessments are often characterised as special, ‘one-off’ projects, and are not integral to organisational culture. Consequently, they are rarely budgeted for (Peberdy, 2008).

Given the high turnover of staff, few incentives – besides personal motivation – exist for individuals to be involved with impact assessments after reassignment from one programme to another. Furthermore, the fact that impact evaluations can demonstrate a range of results – from positive effects, zero effects, or negative effects, can discourage individuals and organisations from effective learning and transparency. Staff and organisations may wish to keep quiet about challenges being faced and work that is not going well, for fear of public scrutiny or criticism (Chapman and Mancini, 2008). A recent ODI scoping study, looking at improving coordination and uptake of impact evaluation, found that only a limited number of impact-evaluation findings – in both development and humanitarian sectors – are published, and these tend to be those containing favourable results (Jones et al., 2008).

Participants at the December 2008 biannual meeting had mixed views about this issue. Some perceived a tension between transparency and ‘keeping the media and donors happy’. They bemoaned the media for equating every aid failure with scandal and donors for commissioning impact assessments simply to gain profile or demonstrate value for money to politicians and the public. Others felt that greater agency transparency leads to increased credibility and therefore better chances of funding from donors who are no longer impressed with over-simplistic ‘success’ stories, and are keen for results. As ALNAP’s research on the media and humanitarian agencies highlights, aid agencies clearly need to be more transparent, but journalists and donors need to recognise that simplistic efforts aimed at transparency and results, respectively, can inhibit openness, learning and improvement (Proudlock and Ramalingam, 2008).

Participants at the biannual meeting identified a range of disincentives generated within their agencies including:

- the complexity of the impact terminology and debate, which deters many from even starting the attempt to assess impact;
- lack of capacity and skills in impact assessment;
- timing of institutional project and budget cycles, which do not always correspond to the necessary timing of impact assessments; and
- absence of impact assessment in research programmes and strategic performance frameworks within organisations, contributing to an overall lack of institutional demand for it.

Furthermore, the presence of ‘perverse incentives’ can lead to impact assessment being done for the wrong reasons. For example, the desire to claim credit for outcomes and impacts can fuel competition and lead individual agencies to downplay the role of others in achieving results (Roche, 2008). As one key informant observed, “A major difficulty [of results-based management] is that UN agencies don’t want to see their work as a contribution, they want gold stars”.

As noted by James Darcy, Director, Humanitarian Policy Group, ODI, during the biannual meeting, a negative consequence of this is that it can lead agencies to work only in areas where they know they can solve problems. Furthermore, policy-

makers and managers have more discretion to pick and choose strategic policy or programme directions when less is known about ‘what works’ (CGD, 2006). As Clark et al note, “Perverse incentives thrive in the absence of information” (2006, p. 230). Where independent consultants carry out impact assessments, explicit and implicit pressures may exist that mitigate against making recommendations to end donor assistance. As one key informant pointed out, evaluators may be reluctant - and find it difficult - to ‘speak truth to power’.

Last but not least, within the humanitarian sector, there are several cultural biases that hinder good-quality impact assessment. These include: the tendency to value action over analysis (ALNAP, 2002); increasing aversion to risk, reflecting growing competition for resources (Watson, 2008); and the tendency, not unique to humanitarian work, to maintain previously held beliefs as common wisdom and neglect new evidence that might contradict these (Roche, 2008). A parallel study on innovations looks at the concept of innovation in more detail, and suggests how innovation can be encouraged in humanitarian work.

Evidence suggests that an understanding of the incentives within humanitarian agencies and across the sector, coupled with determination of agencies to promote positive incentives could lead to more and better impact assessment and, provided findings are used effectively, better results. However, such efforts may be costly in terms of individual agencies’ own time and resources, and may not be funded by donors. One possible solution, the potential of which remains unexplored, is for a sector-wide initiative on impact assessment. The development sector includes several system-wide initiatives to strengthen capacity in impact assessment, such as the Network of Networks on Impact Evaluation (NONIE) and the International Initiative on Impact Evaluation (3IE). However, no such mechanism has developed in the humanitarian sector. Speaking at the December 2008 biannual meeting, Nick York of DFID stated, ‘We must find a way to build systems whereby public goods [such as impact assessments] can be provided more often, because they are under-supplied.’

Key points on capacities and incentives

- 1 There is a lack of individual and organisational capacity to carry out good impact assessment within the humanitarian sector. High staff turnover, lack of a 'learning culture' and inadequate resources are contributing factors.
- 2 Disincentives for effective evaluation, learning and transparency include: the complexity of impact-assessment methodologies; the rigid nature of agency programme and budget cycles; and aversion to the perceived risks of failure.
- 3 Perverse incentives can lead to impact assessment being done for the wrong reasons or in the wrong way. For example, the desire to claim credit for impact can fuel competition and lead individual agencies to downplay the role of others.
- 4 There may be potential for sector-wide initiatives to strengthen capacity in humanitarian impact assessment, and to address some of the disincentives and costs.

2.3 Humanitarian impact assessment in practice

This section (2.3) builds on the framework outlined and explored in the previous section (2.2), to explore experiences and lessons from four different humanitarian initiatives on impact assessment.

Using the structure of the framework, the analysis of the case studies examines:

- the motivations of different stakeholders and the purpose of the assessments;
- how impact was defined and operationalised;
- the methods used to demonstrate, and collect evidence of, impact;

- the institutional contexts within which they were implemented, with a particular focus on capacities and incentives; and
- the subsequent use of the impact evaluation findings.

2.3.1 Overview of the case studies

The case studies were selected through a combination of identifying examples of impact assessments from a first round of research, soliciting further examples from the ALNAP membership, and searching web-based databases (as detailed above near the end of Section 2.4).¹³

In consultation with a cross-network Advisory Group on Impact Assessment, the final four cases were selected because they:

- 1** focused on humanitarian crises
- 2** included an explicit definition of impact
- 3** were ‘researchable’ in that they had available data in the form of reports and key informants willing and available to provide insights and inputs into the process through interviews
- 4** provided useful comparisons between different approaches to key challenges of impact assessment.

The case studies are summarised in Box 2.5.

As will have been noted from the box above, there are considerable differences between the four initiatives. For example, the scope of programmes assessed varies dramatically – ranging from assisting ten pastoralist communities in one valley, to the entire post-tsunami recovery efforts in four countries. The assessments also differ in terms of the stage of the project cycle at which they were undertaken. For example, the IRC evaluation was a prospective study, planned and incorporated into the CDR programme design from the outset. In contrast to this, the FAO and PIA evaluations were retrospective studies of programmes for which the funding had ended. TRIAMS is an ongoing initiative and looks both backwards and forwards. The

Box 2.5

The four case studies

1 Impact evaluation of CDR, Northern Liberia (IRC)

This evaluation, completed in December 2008, assessed the impact of a DFID-funded programme of community-driven reconstruction implemented by the International Rescue Committee (IRC) in post-conflict Northern Liberia between September 2006 and February 2007. In 42 communities in Voinjama and Zorzor districts, Lofa county, the community-driven reconstruction programme featured the use of block grants, elected decision-making institutions, participatory planning and community development projects to improve socio-economic welfare, local governance and community cohesion. The impact evaluation, which was carried out by researchers from Columbia and Stanford Universities, with assistance from the Liberia Institute for Statistics and Geo-Information Services and the National Ex-Combatant Peace-building Initiative, looked at whether the programme achieved its objectives. It represents one of the first uses of the method of randomised study for a community-driven reconstruction programme. As the evaluation was completed only in December 2008, the findings have not yet been fed into policy. The estimated cost of the evaluation is US\$200,000.

2 Participatory impact assessment (PIA) in pastoral communities, Niger

This participatory impact assessment (PIA) assessed the impact of a Pastoralist Survival and Recovery Project (ARVIP), implemented between December 2005 and November 2007 in North Dakoro, Niger. The ARVIP project, was designed to assist ten pastoral communities (3,800 pastoralists) living in the Tarka Valley to recover from the effects of a prolonged drought in 2004, and to enhance their capacity to cope with the effects of future droughts and other shocks by ensuring year-round access to food and income. Carried out by project staff from Lutheran World Relief (LWR) and Contribution à l'Éducation de Base (CEB), and supported by the Feinstein International Center, Tufts University (FIC, Tufts), the PIA focused on changes in household food security and income resulting from the project's re-stocking and animal-feed-bank activities. It also examined how the re-stocking activities might have potentially improved participants' resilience to future droughts and other shocks. The assessment was one component of a broader applied research initiative supported by the Bill & Melinda Gates Foundation, which focused on the development and application of a PIA toolkit.¹⁴ The estimated cost of the evaluation is US\$15,000.¹⁵

CONTINUED

3 Impact study of FAO's emergency programme in DRC, 2005–2007

This impact study, carried out between November 2007 and April 2008, sought to assess the contribution of FAO's emergency and rehabilitation programme to food and livelihood security in the Democratic Republic of Congo. The overall goal of the programme, which began in the late 1990s following an influx of refugees from Rwanda and was extended into a rehabilitation stage during 2003–2007, was to save lives and to protect, strengthen and reconstruct the means of subsistence of vulnerable populations following the war and natural disasters. Specific objectives included rehabilitation of roads and small farming schemes, rebuilding seed banks and tool distribution, strengthening rural production and reintegrating demobilised soldiers in agricultural activities. At village level the impact assessment focused on the extent to which programme activities had strengthened the subsistence activities of vulnerable populations in terms of food security and income. At district and province level, the study looked at the effects of the programme on local organisations and the extent to which programme activities influenced overall vulnerability. The study was one component feeding into a broader evaluation of FAO's entire national portfolio covering the same period and conducted between September 2007 and June 2008. The study cost roughly US\$100,000.

4 Tsunami Recovery Impact Assessment and Monitoring System (TRIAMS)

TRIAMS is an initiative of the governments of Indonesia, the Maldives, Sri Lanka, and Thailand,⁴⁶ with support from the IFRC, WHO and UNDP/BCPR,⁴⁷ to monitor recovery efforts following the 2004 Indian Ocean tsunami, and to assess the overall impact of those efforts. The Global Consortium for Tsunami-Affected Countries endorsed a recovery impact assessment and monitoring system, based on ideas previously put forward by the IFRC. The Global Consortium gave a mandate to IFRC and WHO to develop the system, with the support of the UN Office of the Special Envoy (OSE), to catalyse and harmonise the collective efforts of national agencies, ministries and international and national organisations to monitor tsunami recovery activities, and to assess the impact of what had been achieved so far in the tsunami-affected countries across all recovery sectors. TRIAMS is now in its fourth year and is scheduled to run until 2010. The overall cost of TRIAMS IFRC, WHO and UNDP between January 2006 and June 2009 is estimated to be US\$542,000.⁴⁸

assessments also differ in the extent to which they were seen as an opportunity to field-test tools and methods, rather than to learn about the past, and also in timeframes and methods.

These differences are important here because they limit the possibilities for detailed comparative analysis across the case studies. The section that follows is deliberately exploratory and illustrative in nature, in line with the overall aims of this study. However, it is important to acknowledge that as a result, it may raise more questions than it answers, especially for those with in-depth experience of impact-assessment methodologies.

2.3.2 Understanding and balancing stakeholder interests in impact assessment

One of the key messages from the literature is that clear understanding of why an impact assessment is being done, who the stakeholders are and what specific questions they are seeking to answer, is paramount. Without this understanding it is difficult, or impossible, to establish appropriate indicators, select appropriate methods and generate results that are actually used.

All four initiatives brought together multiple stakeholders, including implementing agencies, donors, academics, consultants, and recipients of aid. It was challenging in all cases to agree a common sense of purpose and specific questions which are valuable and interesting for key stakeholders. A common reason for the difficulty was perceived tension between the separate goals of accountability to donors and learning; in some cases, this caused problems in later stages of the impact assessments.

For example, the Gates Foundation, which commissioned the PIA initiative in Niger, was interested in understanding the impact of its funding for innovative humanitarian programmes. However, Gates executives also perceived a need to develop and field-test a more flexible, participatory approach to impact assessment in the humanitarian sector. The objectives of the initiative therefore included a research and capacity-building component which focused on developing and field-testing PIA methods and tools. This dual goal led to a lack of clarity among implementing partners about the overall purpose of the PIA. According to one key

informant, field personnel were wary of being too open about their practices of monitoring and data collection, in case these were being evaluated.

Moreover, for a number of reasons including limited time and resources, the focus on humanitarian ‘innovations’ and a desire for consistency, the PIA in Dakoro focused on just two aspects of the ARVIP project – re-stocking and animal feed-bank activities – although ARVIP also included many other activities. Consequently, the impact assessment served the interests of the donor and academic researchers more than those of other stakeholders. Indeed, Lutheran World Relief (LWR) commissioned an entirely separate evaluation to meet its learning needs. A perception highlighted in the LWR evaluation is that the interests of the donors and FIC, Tufts were, to some extent, contrary to those of the implementing agencies (Cekan and Hlaibi, 2008).

IRC’s evaluation of its CDR programme in Liberia was the agency’s first venture into randomised impact evaluation, and largely driven by an interest in new and better ways of learning about programme effectiveness and results (Nelson, 2008). The decision to implement an RCT in a post-conflict context came about via IRC’s partnership with the Center on Democracy, Development and Rule of Law at Stanford University. However, there was still a tension between the desire to learn and the need for accountability to IRC’s donors:

‘The most interesting research questions would increase our overall knowledge about key processes... as such they could help us [IRC] to explore whether some of our tacit assumptions and theories underlying programme design were valid. On the other hand, that the evaluation would be done of an IRC programme, and that the donor was very interested in the results, put pressure on staff to not steer too much into building knowledge out of concern that findings would be misconstrued as poor performance’ (Nelson, 2008).

As a large, multi-country, multi-sectoral programme, TRIAMS depends for success on buy-in from as many different agencies as possible. Key stakeholders include: the UN Office of the Special Envoy for Tsunami Recovery, led by former President Bill Clinton, who, according to key informants, sees TRIAMS as a way of promoting accountability to donors and donor publics; the International Federation, which proposed the idea and was interested in a collective system to ensure needs were adequately met; WHO which was interested in building the capacity of participating

governments to monitor the tsunami recovery efforts and make decisions in real time; and the governments of tsunami-affected countries and their relevant counterparts,⁴⁹ which recognised the benefits of using a common language and framework to communicate and work with international donors. Given the programme's broad constituency, TRIAMS has had considerable success in achieving consensus among the four countries most affected by the tsunami (Indonesia, Maldives, Sri Lanka and Thailand) around a core set of recovery output and outcome indicators. Agreeing on and developing these indicators was a major undertaking that lasted more than 18 months.

The FAO Impact Study was carried out to promote accountability to donors to the FAO emergency programme, and to improve the relevance, performance, results and impacts of FAO's work in DRC. According to one key informant:

'these are classic evaluation objectives used pretty much for all evaluations. They embody, and perhaps simplify, a necessary balancing of stakeholder interests. Accountability is important for the higher-ups and learning and improvement is key for the programme managers, implementing partners and beneficiaries.'

FAO attempted to address the needs of both these audiences through distinct 'products':

'Donors and HQ staff get a debriefing and a report posted on our Internet site, but that's only the tip of the evaluation-communication iceberg: there are fuller, longer debriefings with partners and staff to validate the results and communicate them more deeply.'

Together, the case studies highlight the importance of *negotiating* stakeholder ownership of, and demand for, impact evaluations as early as possible in the process of design and implementation.

2.3.3 What is humanitarian impact? Definitions and theories of change

All four initiatives define impact explicitly (Box 2.6). As in Roche's definition of impact discussed in Section 2.2.2, all focused attention on changes in people's lives

Box 2.6

Case study definitions of ‘impact’

IRC The net difference that IRC’s work makes in people’s lives.

PIA Those benefits and changes to people’s livelihoods, as defined by the project participants, and brought about as a direct result of the project.

FAO Positive and/or negative changes induced (more or less directly) by FAO emergency interventions on target groups, their households, organisations, communities or on the environment in which they live.

TRIAMS Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

as well as on attributing responsibility for those changes. Both TRIAMS and FAO explicitly refer to impact as either positive or negative. Apart from TRIAMS, which uses the OECD-DAC definition of impact, none explicitly refer to unintended outcomes. According to those involved, TRIAMS adopted the OECD definition because it was a ‘known and acceptable’ definition among the key players (OSCE, IFRC, WHO), thanks to the previous work of OECD-DAC. This contrasted with the difficulty of achieving agreement between the three agencies in many other areas.

None of the four assessments defined humanitarian impact in terms of changes relative to what would have happened without the programme. By focusing on the perceived ‘benefits’ in people’s lives, the PIA definition fails to recognise impact as potentially negative. While the participatory methods used were designed to capture negative changes in indicators identified, such as resilience to drought, income or food security, participants in the ARVIP project were asked only to identify *benefits* resulting from the project, and so negative impacts may have been excluded. According to FIC, Tufts, one reason for this is that they feared that implementing partners would be less willing to field-test the PIA methodology if the assessment focused on negative impacts. This points to a trade-off between meeting stakeholder needs and methodological rigour. PIA could easily be designed to capture negative impacts, by simply asking affected populations to define these at the outset, although it is uncertain whether they would be willing to openly identify negative impacts.

IRC's definition emerged from a consultative process with staff members about the meaning of impact. When asked, staff members working in different countries said that they saw 'impact' as the changes that can be *attributed* to IRC's work – or the change in outcomes for which IRC is responsible. This is the definition that brought IRC into contact with academics using random assignment methods. However, IRC staff and evaluators struggled to identify the intended impacts of the CDR programme, and to define indicators of these after only two years. The CDR programme focused on governance – promoting democracy, social cooperation and attitudinal change. As one key informant noted, governance is 'an inherently murky concept', for which outcome and impact indicators are far harder to define than they are for programmes on food security or health. To say, 'we would like household incomes to increase' is a lot simpler than saying 'we would like to increase the inclusion of marginalised communities or support local democracy.'

Generic definitions, as listed in Box 2.6, can be important for setting the parameters of impact assessment, and understanding how impact assessment differs from monitoring and other kinds of evaluation. In each case, however, the generic definition only took the impact-assessment effort so far. As explained in Section 2.2.2, generic definitions can provide a starting point for an evaluation, but something much more concrete and context-specific is required to assess the results of specific programmes. Impact assessments also require that the 'theory of change' underlying an intervention is clear and logical. However, as the IRC case study illustrates, this is easier said than done. If a theory of change is flawed, this may be apparent only later in the evaluation process when it comes to collecting and/or analysing data – as the following section demonstrates.

2.3.4 Methodologies for assessing humanitarian impact

Selection and use of indicators

The case studies illustrate the critical need to define appropriate, context-specific impact indicators (and the sources of verification of these), as well as to think carefully about exactly when they are likely to be visible and measurable. The guiding principle of the PIA initiative was to measure impact indicators identified by project participants themselves. Participants in the ARVIP re-stocking activities defined food security in terms of livestock ownership, and said that in the event of

drought they would sell their sheep to buy millet. It was therefore collectively agreed that 'sale of sheep' would be an appropriate impact indicator because, in the event of a drought, these transactions would demonstrate people's increased resilience. The sales would also be well captured in the 'before and after' scoring exercises of food and income sources and expenditure patterns.

However, choice of this indicator was predicated on the assumption that there would be a drought. In 2007, there was a good harvest and participants did not need to sell any sheep, with many preferring to focus on the longer-term objective of converting these assets into cattle. This made it difficult for the evaluation team to quantify the results of the re-stocking exercise in terms of food and income transfers. Consequently, none of the results turned out to be statistically significant, despite the fact that participants felt that they were definitely better off, and more resilient to drought and other shocks (Burns et al., 2008). This example shows that impact is largely unpredictable and therefore difficult to measure with pre-set indicators. Assumptions must be constantly reviewed, and indicators reassessed and adjusted as the context changes, to avoid collecting meaningless data and potentially underestimating the impact of a project.

Adjusting indicators in line with contextual changes was a key component of the discussions at the second TRIAMS regional workshop. Some TRIAMS indicators were more emergency-based and therefore appropriate early on. Indicators of disaster-risk reduction were thought to be important but not appropriate during the early stages of recovery. Participating governments noted that indicators should be adapted as the context changes, but not so much that changes could not be tracked over time.

The FAO study focused on a mixture of output, outcome and impact indicators. Within its seed-multiplication and distribution projects, it measured agricultural production and income levels. These can be viewed as proxy indicators of impact because the study assumes that they are automatically translated into improved household food security and living conditions. As highlighted in the literature review (Section 2.2.2), this may not necessarily be the case. If the project was implemented in an insecure area, it is possible that the harvested crops never reached the intended recipients, or that they were looted by militias shortly after the harvest. However, the evaluation team gathered data on changes in levels of consumption, number and diversification of meals, which went some way towards exploring how the yields and income were used. However, in relation to efforts to supply

agricultural, breeding and fishing kits, the study focused on outputs. The evaluation concluded that beneficiaries were ‘slightly better off than non-beneficiaries, in particular for agricultural tools, small farm animals, agricultural fields, fishing equipment and bicycles, radios and mobile phones’ (Bangui, 2008).

TRIAMS gathered data on both output and outcome indicators as defined by the OECD-DAC (see Box 2.1 in section 2.2.2). ‘Core’ output and outcome indicators identified during consultations with the participating countries and agencies in 2005 were divided into two sets according to the method required to collect the data:

- 1 those that could be monitored through routine information systems (e.g. kilometres of roads restored, number of teachers trained)
- 2 those that would require supplementary household surveys (e.g. percentage of children under five who are wasting; percentage of population without basic sanitation facilities).

The core indicators were not exhaustive; each country expanded or adapted the list according to its monitoring systems and the specificity of the ongoing recovery efforts in the affected areas (Stansberry and Zagaria, 2006).²⁰ Over time, the TRIAMS experience has shown that ‘less is more’ when it comes to ensuring the quality, timeliness and use of core-indicator data.

Confounding factors and the challenge of attribution: selecting appropriate methods and tools

This section looks at how the different case studies address the attribution problem – determining whether any observed changes are caused by a particular intervention (Section 2.2.3). As highlighted above, approaches for addressing attribution can be broadly categorised into two groups: comparative and theory-based. Depending on the purpose, questions, indicators, context and resources available, methods may be qualitative, quantitative or a mixture of both. The case studies used a range of different approaches and methods to establish a causal link between their outputs, outcomes and impacts. As will be shown, each has strengths and weaknesses.

All of the case studies except for TRIAMS adopted a comparative approach in that they attempt to establish either a ‘before and after’ or a ‘with and without’

counterfactual. All except the IRC evaluation also make use of theory-based approaches.

The IRC evaluation was the only assessment which adopted a randomised approach to analysing attribution. By randomly assigning a large sample of communities to the CDR programme, the evaluators could assume with confidence that any difference between the communities who received the programme and those that did not was caused by the programme (rather than by other programmes, events or trends within the communities). Random assignment also ensured that the CDR programme was not given to particular communities because of particular attributes of those communities (i.e. close to road, effective leadership). In other words it avoided selection bias, which reduced the chances of evaluators attributing differences between programme and non-programme communities to the CDR intervention when they were actually due to pre-existing differences.

Despite the randomisation, however, the evaluators did in fact identify two possible differences between the communities that received the programme and those that did not – these occurred by chance as a result of the randomisation process: the first was that the intervention areas were, on the whole, more rural than the control areas. The second possibility was that receiving the CDR programme could itself induce changes affecting outcomes not directly attributable to programme activities. For example, it is possible that once selected to a CDR programme, a community becomes more likely to receive more or less aid from different sources. To examine this possibility, the evaluators collected data on other programmes taking place in the area but found no evidence that selection into the CDR programme resulted in differences in exposure to other programmes (Fearon et al., 2008).

The randomisation process itself presented a number of challenges. Long distances between randomly selected communities presented serious logistical challenges for the programme, especially because some of the more remote communities are inaccessible by road (Nelson, 2008).²⁴ Another potential problem with randomised evaluations is that they may give rise to jealousies, and even conflict, between communities who did and did not receive benefits (Fearon et al., 2008). However, one of the most interesting findings of the IRC evaluation is that the public lotteries were very well received and communities expressed happiness with the process, which was perceived as fair and transparent.

The FAO study also adopted a comparative approach in order to establish a counterfactual. Specifically, it compared the situation before and after the intervention using recall methods, as well as comparing households ‘with’ and ‘without’ the intervention. The assessment team carried out a survey of 1893 households, of which 73 per cent were beneficiaries of FAO aid between 2005 and 2007, and 27 per cent were non-beneficiaries living in the same areas. The latter had the same social and economic characteristics as the former and in that sense represented a control group.

However, both of these approaches have limitations. First, ‘before and after’ recall methods require beneficiaries to remember what happened several months or years previously, so the reliability of data is questionable. Second, comparing beneficiaries and non-beneficiaries living in the same area is problematic as it cannot account for the possibility of beneficiary leakage, or the presence of other systematic differences between the two groups. For example, the non-beneficiaries may not be receiving aid because they are wealthier, or because they are served by other aid programmes. According to key informants from the FAO, sharing of assistance beyond project participants happens quite a lot, and qualitative, open-ended approaches are key to finding out about such unintended effects.

The FAO evaluation team also used a mixture of quantitative and qualitative methods to build a theory of causality. Questionnaires were specifically designed to enable evaluators to explore the multiple possible causes of outcomes. Story-telling methods were also used, involving evaluators sitting down with a handful of targeted recipients and asking open-ended questions about what had happened over the past 18 months. The latter was useful for identifying causal factors not previously considered by the evaluation team.

The PIA study used participatory, comparative scoring and ranking methods to assess project attribution, asking participants to identify and compare all the ‘project’ and ‘non-project’ factors that had contributed to any assessed change. This approach is thought to be particularly useful where the use of a control group is unethical or impractical, as is often the case in humanitarian assistance (Catley et al., 2008). However, the impact assessment team experienced a number of difficulties in attributing improvements in household food security and income to a particular cause or project activity.

Specifically, the evaluation team asked participants to identify any perceived improvements in food security or income, and the reasons for these improvements. All the reasons given were then categorised and tallied to give a score. The most frequently mentioned reasons would score highest, giving a rough representation of their relative importance in comparison to other factors. However, analysis revealed that participants attributed improvements to several factors at once, making the assigned scores less informative. For example, reasons given for improved food security included:

- ‘we get more income and milk from our livestock because of the feed bank and habbanaye’²²
- ‘we can get more income from our livestock because they are now healthier’
- ‘we now have livestock to sell in order to meet our needs’
- ‘we are food-secure because our livestock are healthier as a result of the feed bank and well’.

There is clearly overlap in these responses, with income also representing food security. Part of the final analysis involved trying to unpack and synthesise these responses. With hindsight, the evaluators felt that it would have been better to identify all the factors contributing to improved food security and rank these in order of importance. According to the evaluation report, the decision not to do this was influenced by the limited language skills of the FIC consultant overseeing the assessment, and some of the tools were simplified to accommodate this limitation (Burns et al., 2008). Contextual factors also posed problems. The PIA team struggled to attribute qualitative responses by beneficiaries, as it was unclear whether the positive assessment was due to the project or to the good harvest that year. This is a key example of the limitations of data due to intervening variables.

Neither the FAO nor the PIA initiatives adopted a randomised approach to sampling, and were therefore prone to selection bias. The FAO team originally wanted to carry out the study in North Kivu, but ongoing fighting in this area forced its elimination from the sample frame. The final sample also excluded displaced people and other mobile groups who were almost impossible to track, and thus the sample was not representative of all FAO beneficiaries. It is therefore possible that the selected areas may have had pre-existing features (e.g. greater

accessibility) which would influence the results of the intervention. Similarly, the communities in the PIA study were selected by the implementing agency based on considerations of ethnic diversity, the presence of project participants at the time of the assessment and distance from Dakoro town. The results may therefore be shaped by factors other than the project itself.

Given the scope and scale of TRIAMS, it is not surprising that it did not attempt to establish a counterfactual. Instead, the idea was to monitor post-disaster social changes over time so that these could be compared and addressed by governments and organisations. The tracking of changes in outcome indicators over time, enabling a before-and-after comparison, could be interpreted as a measure of the impact of the recovery efforts. The key problem is that, even where 'before' data are available, these may not be a good proxy for how outcomes would have evolved over the recovery period without the recovery programmes being assessed. In other words, if the situation in particular countries would have deteriorated were it not for recovery assistance, the design would risk underestimating the impact of the recovery efforts. In countries where the recovery capacities of communities and local authorities are fairly strong, it could risk over-estimating the positive causal impact of recovery efforts (after Fearon, 2004).

As shown above, the case studies highlight the strengths and weaknesses of comparative and theory-based approaches to assessing causality and attribution. They also demonstrate that it is not easy to establish a counterfactual, as this involves imagining and collecting evidence on what the situation of targeted aid recipients would have been in the absence of the intervention. The TRIAMS experience also raises the interesting question of what types of counterfactuals are appropriate or possible in large-scale impact assessments in multiple locations and with multiple data-collecting partners and methods.

Quantitative and qualitative methods

The case studies also highlight some of the strengths and weaknesses of quantitative and qualitative methods.

As Table 2.3 shows, a range of different methods was employed across the case studies, but three of the four assessments used a mixture of both qualitative and quantitative methods. This enabled the triangulation of data and proved useful for overcoming the specific limitations of each method.

Table 2.3 Methods used in the case studies: a summary

Case study	Qualitative	Quantitative
IRC impact evaluation of governance activities in Liberia		
	Semi-structured interviews, behavioural goods measurement	Randomised evaluation, 'before' and 'after' household survey
LWR participatory impact assessment (PIA) of support to pastoralists' food security and income in Niger		
	Semi-structured household interviews, focus-group discussions, 'before and after' ranking and scoring	(Not applicable)
Impact study of FAO emergency and rehabilitation work in rural DRC		
	'Before and after' recall methods, story-telling, focus-group discussions, qualitative survey of NGOs	Household survey with 'control' group
TRIAMS (Tsunami Recovery Impact Assessment and Monitoring System): IFRC and WHO programme in Indonesia, Maldives, Sri Lanka and Thailand		
	Encouraged qualitative approaches, including beneficiary perception surveys	Routine data-collection, household surveys

Both the IRC and FAO studies involved quantitative surveys, which were then triangulated against qualitative data, and so usefully illustrate the benefits of mixed approaches. The qualitative data were derived from a community-wide behavioural public-goods 'game'²³ in the case of the IRC, and from a mixture of semi-structured interviews, focus-group discussions, village assemblies and life stories in the case of FAO. Both examples highlight the value of adopting mixed methods for overcoming biases or errors:

According to the IRC impact-evaluation report, the survey data collected by the evaluation team did not necessarily reflect real changes in attitudes and

behaviour. Rather, it was likely that recipient communities had learnt what to say and how to act in order to ‘please outsiders’ (Fearon et al., 2008). The evaluators note that methods designed to capture actual behaviour changes (such as the behavioural goods game) are less prone to such bias. These methods are strengthened when they are undertaken independently – in the IRC case, the behavioural goods game was implemented by an established Liberian NGO. The implementers did not know which communities were targeted by the CDR programme and which were not in order to reduce any bias that the implementers might have in measuring results (Fearon et al., 2008). Similar biases were encountered in the FAO impact study in DRC. According to one evaluator, communities and individuals framed their responses to enhance their chances of receiving more aid.

Equally important was the observation that ‘sometimes the respondents are wrong, even if they aren’t biased’. For example, when surveyed, several beneficiaries of FAO assistance identified improvements in their physical security as an impact of receiving FAO seeds and kits. However, according to one of the evaluators, this was unlikely to be the case. If anything, the FAO had (when assistance had a positive effect on production) increased the likelihood of looting of farm produce and thus, insecurity. While beneficiaries perceived a positive causal link between FAO assistance and improved security, other evidence suggested a negative causal link.

The IRC and FAO examples demonstrate that, where possible, evidence about causal connections from beneficiaries needs to be carefully considered and triangulated with other sources of data. According to one key informant:

‘It’s good practice to read [survey] results with a pinch of salt and ask a few ‘alarm bell’ questions... as a check against beneficiary optimism – the tendency for some beneficiaries to say they’re happy about everything in the project.’

Respondents from the FAO evaluation team also highlighted the importance of qualitative methods for gathering views and suggestions for improvement from targeted aid recipients. They asserted that open-ended questions to beneficiaries (for example about what they did with the aid they received, and what aid agencies might do differently) were more likely to reveal actual uses and effects of aid, as well as possible ways to improve.

The PIA methodology relied solely on qualitative methods – household interviews, focus-group discussions, participatory ranking and scoring methods.²⁴ However, these were applied systematically, enabling the evaluation team to assign a numerical value to qualitative data and ultimately produce statistically representative results. Key informants emphasised the fact that even though the data may be subjective, they were systematically collected and therefore scientifically rigorous. They noted one potential disadvantage: in trying to collect numerically representative data, less time was available for collecting richer qualitative data. Researchers could not go into great depth or do unstructured interviews. Debate continues about whether the resulting information can be called ‘quantitative’ data.

TRIAMS advocated a mixture of qualitative and quantitative data-collection methods, but in practice stakeholders gravitated towards quantitative methods because these were thought to ‘paint a more immediate picture’. Qualitative methods such as key-informant interviews and focus-group discussions were carried out when quantitative data could not be explained. For example, when housing figures in Banda Aceh could not be reconciled, additional qualitative work done with the UN and NGOs, helped to inform the decision by Indonesia’s Reconstruction and Rehabilitation Agency, *Badan Rehabilitasi dan Rekonstruksi* (BRR) to carry out a census (Stansberry and Zagaria, 2007).

Baselines, monitoring and data collection

Findings from the case studies clearly support findings from the literature and interviews in highlighting a common obstacle to good impact assessment (or indeed evaluation): lack of baseline and monitoring data. While a baseline is not essential, it is helpful. It is easier to assess impact effectively if baseline data have been gathered and relevant indicators have been monitored efficiently over time.

The IRC evaluation illustrates good practice in this respect. A significant baseline survey of a representative sample of 1,702 households was carried out before the project started. However, because the survey was carried out before the IRC selected treatment and control groups, only 1,606 of these completed surveys ultimately ended up in the sample frame for selecting treatment and control groups. Two years later, after the conclusion of the CDR programme, a follow-up survey reassessed 1,382 of 1,606 households surveyed earlier, plus 166 new households. These

surveys provided basic information on the demographics of households, their exposure to conflict and detailed information about the pre- and post-intervention status of the three main outcomes CDR programmes are hypothesised to impact: socioeconomic welfare, social cohesion, and individual attitudes about democracy. They were also useful for assessing the extent to which the communities randomly selected for the CDR intervention differed significantly from the 'control' communities.

However, there was a lack of accurate data on the units of the sample frame. IRC needed the number and name of all the communities in the two districts in which IRC works, the number of villages within each larger unit and the number of households within each village. This was especially challenging given the post-conflict context in which a census undertaken in March 2008 was the first since 1984 (Nelson, 2008). Also, the baseline survey was held approximately half a year before the beginning of programming. This is a relatively long time in a post-conflict setting, during which there may have been significant changes affecting community welfare, independent of the CDR intervention. As community selection was randomised, effects due to this delay would not have introduced any bias in the study's methods but may have made the ultimate estimates less precise (Fearon et al., 2007).

By contrast, no baseline data existed for FAO's study in DRC. In fact, there was not even a consolidated list of programme beneficiaries and the evaluation team had to spend time creating such a list, together with FAO offices and local NGOs. This is a common problem, which the FAO evaluation team addressed by constructing a baseline retrospectively using 'before and after' recall methods. A limitation of this approach is recall bias, which would become more pronounced over time. This method is therefore more suitable where short-term impact is expected, and less so for longer-term projects.

Under the original proposal, FIC, Tufts were to support the development of the project M&E plans, and baseline surveys in a way *that would capture impact*. However the grant for the research component was only approved once the agency M&E plans had already been developed and most of the baseline surveys had been done. Consequently, the PIA initiative also developed a retrospective baseline using comparative ranking and scoring methods,²⁵ with separate scores for 'before', 'during' and 'after' the project, which could then be compared. Definitions of 'before', 'after' or 'during' were obtained using time-lines which provided a useful reference

for establishing agreement between the assessment team and the affected population on different points in time. The weakness of scoring exercises, however, is that they measure only relative change. For example, they could be used to measure how much livestock production had increased relative to how much crop production had increased, but would not provide an actual figure for the increase.

By promoting the collection of data against a core set of output and outcome indicators over time, TRIAMS has the potential to provide both pre- and post-emergency baseline data for the tsunami recovery efforts, as well as to build national capacities for future disaster responses. However, TRIAMS has experienced a number of challenges with regard to data gathering, as follows.

- Data were often lacking, for example in Aceh due to years of conflict and lack of appropriate institutions or routine systems.
- Data collection was not timely. As mentioned above, this is especially important in humanitarian situations, where the speed of the planning cycle is accelerated. Moreover, opportunities to construct post-tsunami baselines were missed because it was wrongly assumed that baselines could not be established once recovery efforts had started, and even though not all communities had been reached by core interventions.
- Data collection, management and analysis were often weak; existing data are scattered across different record and documentation systems in distinct agencies and at multiple levels. For example, routine health data were available in tsunami-affected districts in Sri Lanka and Indonesia, but were never collated or analysed and therefore not used. Many stakeholders were simply unaware of the data, and others did not know how to use them. At sub-district level, TRIAMS experienced difficulties in ensuring systematic and standardised approaches to data collection.

In promoting the use of routine monitoring data to influence planning and decision-making, TRIAMS emphasised the need for sound monitoring systems and multi-stakeholder meetings to discuss results (including identifying gaps and over-commitments) and joint planning. While not especially innovative, this approach has been under-emphasised in emergency and recovery operations. TRIAMS also developed a framework for collecting data on recovery:

- vital needs (e.g. water, food and shelter)
- access to basic services (e.g. healthcare, education, water and sanitation)
- livelihoods and economic security (formal and informal income generation)
- infrastructure (public works and social infrastructure) (Stansberry and Zagaria, 2006b, 2007).

This information framework, which helped to overcome a lack of needs assessment and planning, can be seen as the real innovation of TRIAMS. It represents an ambitious attempt to contribute to, and advocate for, improved data quality and preparedness in relief and recovery efforts, providing a useful framework for tracking a number of common recovery indicators over time.

Timing of the impact assessment

Timing was a constraint on all four initiatives, except for the IRC evaluation, which was built into the CDR programme from the very beginning. Specifically, timing had a critical effect on the selection and measurement of indicators. Because the FAO study was carried out some time after the start of the programme, the evaluation team had difficulty measuring some of its chosen proxy indicators of impact. Some of the short-term effects such as cultivated area or amount of production sold were no longer present or well remembered.

Given the importance of TRIAMS for facilitating re-planning of tsunami recovery interventions, data should ideally be collected quickly and regularly, preferably on an annual basis. One key informant observed, ‘data needs to be available when the critical mass of interest is primed. It can’t come a year later’. Yet many of the affected countries conduct their socio-economic surveys at much less frequent intervals, which meant that data were often not available to inform decision-making.

Conversely, the PIA initiative was undertaken just 18 months after the project began, which proved to be a significant limitation resulting in the underestimation of project impact. Project beneficiaries defined food security in terms of livestock ownership, and ultimately wanted to breed enough sheep to enable the purchase of a cow.²⁶ Local people identified potential income and food benefits from owning a cow, in

terms of the sale and consumption of milk products. However, these benefits had not yet had time to materialise when the impact assessment was carried out. Therefore, the impact in terms of potential food and cash transfers could be measured only hypothetically (Burns et al., 2008). This led the evaluation team to focus on proxy indicators of impact, such as the number of sheep born from the restocking exercise, rather than the ultimate ownership of a cow.

2.3.5 Improved analysis through engagement with affected populations

Research findings presented above (Section 2.2.4) highlighted the importance of engaging with beneficiaries in improving analysis and impact assessment. However, of all four case studies, only the PIA initiative engaged with beneficiaries beyond their role as survey or interview respondents, by involving participants in the design of the evaluation. In none of the case studies explored here did beneficiaries play a role in the analysis and interpretation of evaluation findings.

Beneficiaries of the ARVIP programme played a role in defining the impact indicators, which shaped the research questions and the selection of methods and tools. In this way, the evaluation attempted to balance the requirements of learning and accountability to both donors and affected populations. The use of community-defined indicators went some way towards meeting the demands of project participants in that they were given the opportunity to assess the project themselves, and, according to one key informant, were ‘given a vote’. In the other three cases, participation of recipients was limited to survey responses, focus-group discussions or village assemblies and, in the case of the IRC evaluation, behavioural games.

A key lesson across all the case studies is that understanding the livelihoods context of recipient populations is an important part of analysing and interpreting data. For example, in the PIA study in Niger most of the potential food and income benefits from the project were tied up in livestock assets. An understanding of the livelihoods context allowed the evaluators to interpret the fact that livestock had *not* been converted into income or food as an indicator of both short- and longer-term food and financial security.

Key informants mentioned that the FAO evaluation was ‘frontloaded’, with much effort invested at the start. According to one respondent:

‘there is a need to strengthen the analysis and interpretation of findings... You spend a lot of time designing the questionnaire and doing the survey and then at the end of the process there is very little time to really analyse and use the information, including doing statistical analyses.’

Conversely, the IRC evaluation relied solely on rigorous statistical analysis, which limits the scope for participation in judging impacts. Interestingly, the evaluators noted that looking for statistically significant differences in specific impact indicators may not be the best way to analyse the overall effects of the programme. For an overall assessment, they emphasised the need to move beyond analysis of specific outcomes, and towards understanding whether there are larger patterns across impacts, and whether these patterns would have arisen if the programme had not been effective. Limited data analysis and use continues to be one of the main challenges facing TRIAMS. To date, little thought has been put into how information can be turned into knowledge and used for planning. Key informants suggested that, rather than inefficiently collecting information for information’s sake, stakeholders should better understand and use the data generated.

While the literature highlights the potential of impact assessments as a resource for affected populations, there was little evidence of this from the case studies. Dissemination and use largely focused on upward accountability, especially to donors. One informant noted:

‘the FAO study did not provide FAO evaluators with new or surprising information... Still, it provided credibility vis-à-vis the donors and programme managers. It also supported the conclusions of the broader evaluation, of which it was a component.’

The main exception was the PIA, in which findings were shared with project participants to give them the opportunity to disagree, validate or add additional information. FAO undertook one debriefing with a limited number of beneficiaries. An important challenge identified in discussions with key informants concerned the representativeness of community members invited to debriefings, and the need to ensure that different perspectives are included.

2.3.6 Capacities and incentives for improved impact assessment

Capacity to carry out impact assessments

According to most key informants, capacities for impact assessment are limited. In particular, the PIA initiative, IRC evaluation and TRIAMS all highlighted a need to build capacity not just for data collection but also for data analysis and interpretation. Impact assessment was described by some as ‘frontloaded’, meaning evaluators often spend a lot of time designing questionnaires and carrying out surveys and less time analysing and using the information.

As noted in an LWR evaluation report:

‘the main benefits that all of the evaluators... brought to ARVIP programme staff was not how to collect data – they had already received ample training from CARE – it was how to interpret it, analyse it’.

Similarly, in the CDR evaluation, IRC staff reported that collecting data was not the problem. Rather, the challenge was that it was ‘unclear if field staff really understood *why* they were being asked to collect different kinds of data’. TRIAMS has made some progress in this area by supporting national institutions in tsunami-affected countries to collect and use data, building an evaluation culture in the process whereby data are shared more transparently and decisions are evidence-based. However, data use is still a major weakness.

There was also a sense, among key informants, that the skill set needed for conducting impact assessments is different to that typically needed or used by organisations on a day to day basis in the field. As such people with these skills are often brought in as many organisations don’t have them at hand and many do not need them on an ongoing basis.

Key informants had mixed views about who should be employed to carry out impact assessments. Some felt that unlike monitoring and other types of evaluation, impact assessment is a specialism which requires external ‘experts’ possessing statistical analysis and other kinds of research skills to be brought in from ‘outside’ agencies. Others felt that impact assessment should be institutionalised within aid agencies and where necessary, the complexity of the methodology should be reduced to be

more in line with the existing capacity of the sector. One of the objectives behind developing the PIA guide was to try and transfer impact assessment skills to implementing agencies, and key informants involved remarked upon how quickly agency staff picked up the approach and methods.

All impact assessments require people who can speak the appropriate language(s) for talking to project participants, and for facilitating and structuring interviews and discussions. While these capacities do not always exist, all key informants were in favour of using national academics, researchers and enumerators wherever possible.

Long-term partnerships between academics, donors, governments, practitioners or implementing partners, and targeted recipients were widely perceived to be an essential ingredient for strengthening capacity for impact assessment. In addition to providing skills, partnerships between academics and NGOs were also identified as lending greater credibility to an evaluation, partly because external evaluators can operate with greater independence in the field. The time horizon of academic researchers is often longer than that of humanitarian programming staff, and can provide an important degree of continuity. In the IRC evaluation in Liberia, the research team outlived four country directors and two programme directors.

An important implication of TRIAMS is that capacities and resources are, arguably, best dealt with by working collectively across agencies and in partnership with national actors. Building on national and local capacities for monitoring and impact assessment is a key principle of TRIAMS, the underlying logic being that this will lead to more accountable and sustainable recovery efforts. One key lesson learned by TRIAMS is the importance of working through existing institutional structures. In this initiative, data collection and analysis have been found to be better in the health sector where there is a greater degree of collegiality, and a set of shared principles and procedures.

Capacity to use impact assessment findings

The use of findings from impact assessments was mixed across the case studies. The sheer scale of TRIAMS meant that it was difficult for staff members to capture all the different ways in which outputs have been used by different stakeholders, but TRIAMS is thought to have directly influenced a number of policy changes in

tsunami-affected countries. TRIAMS has also led to ‘process use’, which refers to the fact that the evaluation and learning capacities of the individuals and organisations participating in TRIAMS were strengthened through the process itself. The PIA findings were primarily used to develop a guide on impact measurement for humanitarian livelihood interventions. The purpose of the guide was to demonstrate that, despite methodological constraints, impact can be measured in a meaningful and systematic way using community-defined indicators. In this sense, learning from the assessment has contributed to improving capacities.

Key informants from the IRC evaluation team stated that it was too early to tell how the results would be used as the study was completed only in December 2008, and huge amounts of data were still being analysed. However, the evaluation had clearly contributed to process use, and involvement in the design of the evaluation had challenged IRC staff thinking about CDR. One of the weaknesses of the evaluation, according to programme staff, is that it was a highly academic exercise, which international and field personnel struggled to understand. Key informants highlighted the need for improved communication of evaluation findings to programme staff, taking into account both knowledge and language issues among different audiences. Informants were unsure whether learning from the CDR evaluation would be applied to their internal processes of monitoring and evaluation. They highlighted reporting frameworks imposed by donors as a major constraint here:

‘If you are an NGO, and you want to do something innovative, you can set up the systems, but in the end you depend on the donor to fund it.’

FAO’s impact study was perceived by some to have led the FAO to challenge its own ideas – conceptual use – and enabled greater confidence in the overall communication of the evaluation findings, especially in interactions with donors. The quantitative nature of the data was seen as carrying more weight in this regard. Moreover, a few months after the evaluation, the FAO office in Kinshasa employed the impact-assessment consultants to do another survey, looking in more depth at more specific issues of partnerships with different types of NGOs.

Incentives for humanitarian impact assessment

Improving impact assessment and accountability for results requires a clear articulation of how agencies can benefit from assessing their impacts. The case

studies suggest that there is still some way to go in this area. Key informants were unanimous in thinking that one of the strongest incentives for humanitarian agencies to carry out impact assessments would be a donor requirement. As one respondent remarked, 'Of course, we are interested in learning, but if donors ask for it our efforts will be accelerated.' Other incentives included improved learning, better designed policies and projects and, from an agency perspective, securing funding. However, key informants were quick to point out the risk that results may be negative and this might cause agencies to lose funding. The potential for agencies to be penalised for poor or negative results was perceived as an important disincentive to learning.

The complexity and cost of carrying out impact assessments were perceived as disincentives. Many key informants emphasised that sufficient time and resources are often lacking. In particular, high costs were associated with large scale surveys and reliance on external consultants or researchers. However, it was also observed that the costs (both financial and social) of not doing impact assessments and continuing to implement poorly designed programmes can be even higher. Good, ongoing monitoring systems which are not too cumbersome were suggested as a way of reducing costs while contributing to meaningful impact analysis.

Some highlighted the importance of asking, 'when is it *worth* doing impact assessments?' For example, one key informant described the FAO study as 'very costly for what we got', but added that assessing and being accountable for outcomes and impacts is a donor-driven imperative for the FAO evaluation department.

Key informants emphasised that incentives for academic researchers may be different from those of donors or programme staff. For example, researchers may be less interested in whether a specific programme worked or not, and more interested in whether an overall programming approach works or not – they may have more focus on lessons that can be drawn for other projects and knowledge that can be gathered about aid processes more generally. This concern with 'external validity' and the desire to link particular interventions to general findings may well affect the type of projects academic researchers will be motivated to work on.

Other respondents drew attention to disincentives for evaluators, especially when the assessment result is negative or there is no discernible impact. One key informant observed, 'the temptation to not report (or to report a 'cleaned-up' version) can be very

strong'. All of the organisations that commissioned the impact assessments in the case studies should be applauded for having the courage to do this, and publish the results. There was no guarantee that the results would be positive, which could be detrimental in terms of both global profile and local standing.

2.4 Conclusions and recommendations

On the basis of the literature and case studies reviewed above, the overall impression gained is that despite considerable challenges, humanitarian impact assessment is not only desirable but possible. An exploration of the lessons learned and practical experience gained from four recent humanitarian impact assessment initiatives shows that, although each has specific strengths and weakness, questions about the causal impact of humanitarian interventions can be addressed in different contexts. However, in summarising current debates and reflections on how to improve humanitarian impact assessment this review has also identified a number of significant challenges in five key areas. Together, these areas form a suggested conceptual framework which could be used as an evaluation tool and as a starting point for developing and improving impact assessment in the humanitarian sector, and possibly in other sectors:

- 1 Understanding and balancing stakeholder interests: who wants humanitarian impact assessment and why?
- 2 Understanding and defining humanitarian impact
- 3 Methodological approaches and challenges: indicators, attribution, baselines, monitoring, time and timing
- 4 The importance of engaging local actors and affected populations
- 5 Capacities and incentives for improved humanitarian impact assessment

The findings from the case study research support and reflect the five-part framework. They highlight the importance of negotiating stakeholder ownership of impact assessments and agreeing a common purpose, including explicitly addressing tensions between accountability and learning – generic definitions of impact can usefully serve as a common starting point. They also underscore the need for clear, context specific understandings of impact and logical ‘theories of change’ to direct the choice of indicators and assist movement beyond outputs towards outcomes and impacts.

While the four initiatives differ considerably in terms of objectives, scope and methodology, combined they offer some common lessons: namely that timing is a critical factor affecting the accuracy of judgements about impact; that a lack of accurate data at all stages of project and evaluation cycles, and over time, is a key constraint – one which could perhaps be best addressed collectively by the sector; that a variety of approaches, methods and tools exist for analysing attribution in humanitarian contexts, albeit with varying degrees of scientific rigour; that a mixture of quantitative and qualitative methods is useful for overcoming potential biases which may threaten the validity of impact assessment results; that an understanding of context is crucial for analysing and interpreting impact assessment findings, and that this cannot be achieved without the involvement of local actors and affected people themselves who know the context and are often best placed to assess how their lives have changed.

Two of the most intractable issues highlighted by both the literature review and case studies are the risk aversion inherent in current approaches to evaluation, and the lack of a ‘learning culture’ in the humanitarian sector. The findings suggest that the future of humanitarian impact assessment lies in linking different partners across the sector, and introducing new actors and perspectives. Effective, long term partnerships between donors and recipient governments, implementing agencies, academics, evaluators and recipient communities are key to ensuring sector-wide and institutional sustainability of impact assessment. This in turn can promote organisational learning and help build capacities for doing impact assessments, as well as mitigate the considerable costs involved. It is also important to ensure that the debate is not dominated by any one actor or group of actors, or by the proponents of particular methodologies.

Perhaps most challenging of all will be to empower recipients of humanitarian aid to have a greater influence in assessments of aid results, not only as sources of information but as key actors in the design, analysis and use of impact assessments. In doing so, the sector will need to find new ways of working to minimise the possible biases identified in the case study research.

If all this can be achieved, there is cause to believe that impact assessments can make a useful contribution to improving humanitarian performance. It has been suggested above that improved impact assessment could contribute to more and better evidence of what works, more robust assessments of needs, greater coherence of data across the project cycle, sustainable partnerships between a wide range of humanitarian actors, better use of scarce resources and greatly improved participation of affected populations in humanitarian aid. This is consistent with ALNAP's survey findings, in which almost 90 per cent of all respondents felt that impact assessment will play a significant role in improving performance in the humanitarian sector over the next five years (ALNAP, 2009).

The final point is a note of caution: the desire to prove impact, as this study has illustrated, is a complex one, and one with many different underlying motivations.

If, as impact assessment gains more prominence and attention and resources, it is pursued first and foremost for narrow institutional or political purposes, the likelihood is that it will become yet another instrument that has failed to deliver on its early promise to improve humanitarian performance. The desire to *prove impact*, if implemented in a narrow and self-interested way, may lead to never in fact *improving the real impacts* of humanitarian work.

There is therefore a need to ensure the ongoing debate on impact – and the actions and practices that result from it – keep at their core the notion of an approach to impact assessment that is shaped by humanitarian principles first and foremost. At the heart of this ongoing effort should be a principled and pragmatic approach to realising the potential of impact assessment to contribute to improving relief efforts and delivering better assistance to those affected by emergencies around the world.

2.4.2 Recommendations

The humanitarian sector should develop and institutionalise *sustainable approaches to impact assessment*. Learning that leads to improvements and innovations in how humanitarians operate should be the primary driver, as this will both ensure the institutional sustainability of impact assessment, and mitigate the considerable costs of doing impact assessments. The following recommendations are grouped into five sections, according to the framework detailed above, and are presented for consideration by the ALNAP membership and the wider humanitarian sector.

1 **Balancing stakeholder interests shaping the scope and purpose of impact assessments**

- Identify appropriate stakeholder analysis tools for use in discussions of impact assessment, which help to make interests explicit and identify common ground.
- Develop a clear understanding of the key elements of impact assessments for use in stakeholder discussions. This could be done with a ‘pro forma’ as developed by ALNAP for standard evaluations of humanitarian action.
- Develop a list of impact-related questions of interest to different stakeholders across the humanitarian sector.

2 **Defining humanitarian impact and developing theories of change**

- Initiate a cross-agency discussion on the feasibility and desirability of a clear definition of humanitarian impacts and outcomes.
- Develop a menu of theories of change for use in the design phase of impact assessments, explaining the different ways in which a particular programme might be expected to achieve its objectives.

3 **Developing methodologies and approaches**

- Develop relationships with **academic partners and other experts in the field** to design and deliver a toolkit outlining the key methods of impact

assessment for use in the humanitarian sector. This could include practical examples of mixed-method approaches.

- Work with 3IE to explore possibilities for using random assignment in the humanitarian sector. Experts should be engaged to help the sector think creatively and offer a range of options appropriate to the ethical values of humanitarianism.
- Develop a shared database of impact indicators that could potentially be used in humanitarian evaluations. Given that indicators should change according to the specific context, it is important for this database to be marketed not as a universal or exhaustive list but as a useful point of reference. It would not replace the important thinking that should go into the selection of indicators for every evaluation. The ALNAP Performance Indicators Interest Group and the work of Sphere on indicators of minimum standards could be useful here.
- Consider impact within the many ongoing initiatives to improve collection and standardisation of baseline data before an emergency, needs assessments and information provision.²⁷
- Undertake further research on the mix of impact-assessment methods most appropriate in the different emergency phases of relief, recovery and reconstruction.

4 Improving analysis and interpretation of findings

- Ensure the views of affected people are centre-stage to ensure the credibility of impact assessments
- Actively promote the use of livelihood approaches as a framework for analysis, because good livelihood analysis can help to identify the short-, medium- and long-term effects of a programme.
- Invest in and build long-term, national and international partnerships for impact assessment between affected populations, academics, donors, governments, civil society and the private sector in order to build capacity in analysis and interpretation.

5 Building capacities and strengthening “good” incentives

- Review existing programming and funding approaches across the sector in terms of how they currently enable or inhibit effective and timely impact assessments.
- Work towards improved project and programme, organisational and sector-wide performance frameworks which explicitly define impact and embed impact orientation in all stages of the project cycle.
- Consider how donors, agencies and the sector as a whole can better reward individuals and organisations for doing effective impact assessments. It may be useful to incorporate ‘awareness of contribution to impact’ into team and individual appraisals or reward staff for participating in impact evaluations, even after reassignment.
- Funding research on the incentives that emerge from particular organisational and evaluation structures and systems and provide training for staff on this issue.
- Invest in mechanisms to ensure effective dissemination and utilisation of evaluation findings. Agencies need systematic processes to use the findings, plus the necessary resources and authority.
- Invest in and build long-term ‘learning partnerships’ between academics, donors, governments, civil society and the private sector in order to share and build knowledge. Ongoing training in data collection and analysis, through the support of donors, should ensure that capacities are sustained over time.
- Explore the scope for a cross-sector initiative on impact assessments, to strengthen sectoral capacity, and to agree key areas of change for which system-wide indicators could be collectively established.

Notes

- 1** The 24th ALNAP biannual meeting was held in Berlin on the 2nd and 3rd December and was hosted by the Ministry of Foreign Affairs, Germany. Its goal was to develop awareness about impact assessment while recognizing its limits. For further information see the meeting report at: http://www.alnap.org/meetings/pdfs/24_meeting_report.pdf
- 2** A number of journals and web-based evaluation databases were explored, including the ALNAP ERD, Disasters Journal and Journal of Humanitarian Action.
- 3** Examples of each of these are provided in the logframe/table in Hofmann et al, 2004.
- 4** Interestingly, OECD/DAC has changed its definition over time. In its Guidance for Evaluation of Humanitarian Assistance in Complex Emergencies in Evaluation and Aid Effectiveness Series no. 1, impact is defined as ‘The wider effects of the project – social, economic, technical, environmental – on individuals, gender and age-groups, communities, and institutions. Impacts can be immediate and long-range, intended and unintended, positive and negative, macro (sector) and micro (household)’ (see: <http://www.oecd.org/dataoecd/9/50/2667294.pdf>). This is also the definition used in the ALNAP EHA Guide (Beck, 2006): http://www.alnap.org/publications/eha_dac/index.htm.
- 5** Just as Roche’s definition uses the term ‘impact’ to conflate outcomes and impacts (Roche, 2000), Patton (1997) uses the term ‘outcomes’ to do the same.
- 6** While some argue that theory-based methods seek to attribute impacts without a counterfactual, others assert that there is a counterfactual but not one which need be explicit.
- 7** Qualitative methods may be required to establish a system’s level of openness.
- 8** <http://www.cemcentre.org/Documents/CEM%20Extra/EBE/EBE1999/Kevin%20Williams.pdf>.
- 9** Some take the view that theory-based approaches also attempt to establish a counterfactual, just not an explicit one (White, 2009).
- 10** This effect is known as the Hawthorne effect – an experimental effect in the direction expected but not for the reason expected; i.e. a significant positive effect that turns out to have no causal basis in the theoretical motivation for the intervention, but is apparently due to the effect on the participants of knowing themselves to be studied in connection with the outcomes measured. For further information see: <http://www.psy.gla.ac.uk/~steve/hawth.html#Hawthorne%20overall>.
- 11** The remaining comments were unclear.
- 12** The PIA Guide is available at: <https://wikis.uit.tufts.edu/confluence/display/FIC/Participatory+Impact+Assessment—+a+Guide+for+Practitioners>.
- 13** Initially, 11 impact-related initiatives were identified. In addition to the four included in Box IA.5, these were: Beneficiary Perception Surveys, Fritz Institute; Emergency Capacity Building project; HAP-I Standard 2007; Health and Nutrition Tracking Service; Listening Project, Collaborative for Development Action; Quality Compass, Groupe URD; and the Standardised Monitoring and Assessment of Relief and Transitions.
- 14** The PIA case study explored here is not necessarily representative of the PIA approach. For example, most other PIAs carried out by Tufts used a mixture of quantitative and participatory methods. Furthermore, in all the other Gates-funded PIAs, the indicators used

remained constant and proved useful in measuring impact.

- 15** This amount would cover a local lead researcher (assuming they already have PIA skills), transport and remuneration and per diems for 4 data collectors. It does not cover external consultants
- 16** India was initially included but dropped out after the first regional workshop in March 2006.
- 17** UNDP/BCPR (Bureau for Crisis Prevention and Recovery) was initially reluctant but ultimately committed to developing TRIAMS indicators into a more generic post-disaster monitoring framework.
- 18** This figure includes travel, consultancy and technical support to governments, annual or bi-annual country level workshops and annual regional workshops.
- 19** BRR in Indonesia, RADA in Sri Lanka, Ministry of Planning and National Development in Maldives, Ministry of Interior in Thailand.
- 20** For a more comprehensive list of TRIAMS core output and outcome indicators see the 2nd TRIAMS Workshop Report (Stansberry and Zagaria, 2007) available at: http://www.who.int/hac/crises/international/asia_tsunami/triams/en/index.html.
- 21** Many of these issues can be addressed when defining the initial population: places that are impossible to reach or do not require the programme can be ruled out before the randomisation process (Fearon et al., 2008).
- 22** ‘Habbanaye’ refers to the traditional targeting mechanism used by the project’s re-stocking activities.
- 23** The use of the term “game” here follows from the fact that the approach employed draws on “game theory”, a branch of economics that examines the behaviour of individuals in strategic settings (Fearon et al., 2008).
- 24** See the PIA guide, at: <https://wikis.uit.tufts.edu/confluence/display/FIC/Participatory+Impact+Assessment—a+Guide+for+Practitioners>.
- 25** Participatory ranking and scoring methods require informants to assess the relative importance of different items or indicators. Ranking usually involves placing items in order of importance (1st, 2nd, 3rd, etc). Scoring methods assign a value or a score to each item, often using seeds or stones as counters. For further information, see Catley et al (2008) Participatory Impact Assessment: A guide for Practitioners, available at: https://wikis.uit.tufts.edu/confluence/download/attachments/19924843/Part_Impact_10_24_08V2.pdf?version=1.
- 26** Focus-group participants in Marafa village estimated that the sale of four adult sheep would be necessary to buy a two-year-old calf – such a calf might conceive and start lactating at three or four years old.
- 27** For example, WFP’s SENAC project on needs assessment, and the ACE project on the Humanitarian Dashboard on information provision.

Bibliography

- ALNAP (2002)** 'Humanitarian Action: Improving Performance through Improved Learning' in *ALNAP Annual Review*. London: ODI. Available at: <http://www.alnap.org/publications/rha.htm> (accessed 26 May 2009).
- ALNAP (2003a)** "Synthesis of Findings of 2001–2002 Evaluation Reports" in *Humanitarian Action: Improving Monitoring to Enhance Accountability and Learning*. ALNAP Annual Review. London: ODI. Available at: <http://www.alnap.org/publications/rha.htm> (accessed 26 May 2009).
- ALNAP (2003b)** "Strengthening monitoring in the humanitarian sector" in *Humanitarian Action: Improving Monitoring to Enhance Accountability and Learning*. ALNAP Annual Review. London: ODI. Available at: <http://www.alnap.org/publications/rha.htm> (accessed 26 May 2009).
- ALNAP (2008)** ALNAP 24th Biannual Meeting Report. London: ODI. Available at: <http://www.alnap.org/meetings/24.htm> (accessed 26 May 2009).
- ALNAP (2009)** *Impact Assessment Survey*. London: ODI
- Bangui, C (2008)** *Etude D'Impacts du Programme de la FAO Urgence en République Démocratique du Congo*. Rome: FAO.
- Barnett, M (2005)** *Humanitarianism Transformed*. University of Minnesota. Available at: <http://www.ssc.upenn.edu/centers/browncip/2005.06.Papers/Barnett.pdf> (accessed 26 May 2009).
- Beck, T (2006)** *Evaluating Humanitarian Action Using the OECD-DAC Criteria: An ALNAP guide for Humanitarian Agencies*. London: ODI. Available at: http://www.alnap.org/publications/eha_dac/index.htm (accessed 26 May 2009).
- Bolton, P, et al. (2007)** 'Expanding the Scope of Humanitarian Program Evaluation' in *Prehospital Disaster Medicine*, Vol 22, No 5, pp. 390–395. Available at: <http://pdm.medicine.wisc.edu/22-5%20PDFs/bolton.pdf> (accessed 26 May 2009).
- Borton, J (2008)** *How feasible is it to monitor and report on the overall performance of the humanitarian system?*: Draft overview paper of the ALNAP Humanitarian Performance Project. Unpublished manuscript.
- Burns, J, et al. (2008)** Impact Assessment of the Pastoralist Survival and Recovery Project Dakoro, Niger. Medford, MA: Feinstein International Center, Tufts University.
- Catley, A, et al. (2008)** *Participatory Impact Assessment: A Guide for Practitioners*. Medford, MA: Feinstein International Center, Tufts University. Available at: <http://wikis.uit.tufts.edu/confluence/display/FIC/Participatory+Impact+Assessment> (accessed 26 May 2009).
- CDA Listening Project (2008)** *International Assistance as a Delivery System*
- The CDA Listening Project Issue Paper, September 2008**. Available at: http://www.interaction.org/files.cgi/6604_ListeningProject_IssuePaper_DeliverySystem.pdf (accessed 20 May 2009).
- Cekan, J and JA Hlaibi (2008)** *Evaluation report of Pastoralist Survival and Recovery of ARVIP (Appui a la Réhabilitation des Systèmes de vie Pastoraux) and of CEB (Contribution a l'Education de Base), the local partner NGO*. Lutheran World Relief.
- CGD (Center for Global Development) (2006)** *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Report of the Evaluation Gap Working Group. Available at: <http://www.cgdev.org/content/publications/detail/7973> (accessed 26 May 2009).

Chapman, J, and A Mancini (2008)

Impact Assessment: Drivers, Dilemmas and Deliberations. Prepared for Sightsavers International. Available at: http://www.bond.org.uk/data/files/resources/154/sightsaversimpactassessment_june08.pdf (accessed 8 June 2009).

Clarke, P, and B Ramalingam (2008)

'Organisational Change in the Humanitarian Sector' in *ALNAP 7th Review of Humanitarian Action*. London: ODI. Available at: <http://www.alnap.org/publications/7RHA/Ch2.pdf> (accessed 26 May 2009).

Crisp, J (2004)

'Thinking Outside the Box: Evaluation and Humanitarian Action' in *Forced Migration Review 8: 4–7*. Available at: <http://www.fmreview.org/FMRpdfs/FMR08/fmr8.2.pdf> (accessed 26 May 2009).

Darcy, J and CA Hofmann (2003)

According to need? Needs assessment and decision-making in the humanitarian sector HPG Report 15, London: ODI Available at: <http://www.odi.org.uk/resources/download/239.pdf> (accessed 26 May 2009).

Emergency Capacity Building Project

(ECB) (2007) *Impact Measurement and Accountability in Emergencies: The Good Enough Guide*. Oxfam Publishing.

EES (European Evaluation Society)

(2008) *ESS Statement: The Importance of a Methodologically Diverse Approach to Impact Evaluation – Specifically with Respect to Development Aid and Development Interventions*. EES, December 2008.

Fearon, J (2004) 'Measuring Humanitarian Impact'. Unpublished manuscript.

Fearon, J, et al. (2008) *Community Driven Reconstruction in Lofa County: Impact Assessment*. Unpublished manuscript.

Fearon, J, et al. (2007) *Community-Driven Reconstruction in Lofa County: Baseline Survey Preliminary Report*. Unpublished manuscript.

Forss, K, and S Bandstein (2008)

Evidence-based Evaluation of Development Cooperation: Possible? Feasible? Desirable? NONIE Working Paper 8. Available at: http://www.worldbank.org/ieg/nonie/docs/WP8_Forss_Bandstein.pdf (accessed 26 May 2009).

Harrell-Bond, B (1986)

Imposing Aid: Emergency Assistance to Refugees. Oxford: Oxford University Press, pp. xi–xii.

Hofmann, CA, et al. (2004)

'Measuring the Impact of Humanitarian Aid' HPG Research Briefing 15, London: ODI Available at: <http://www.odi.org.uk/hpg/papers/hpgbrief15.pdf> (accessed 26 May 2009).

Hospes, O (2008)

'Evaluation Evolution? Three Approaches to Evaluation' in *The Broker*, issue 8, June. Available at: <http://www.thebrokeronline.eu/en/articles/Evaluation-evolution> (accessed 8 June 2009).

Jones, H (2009)

The 'gold standard' is not a silver bullet for evaluation Opinion Piece. London: ODI. Available at: <http://www.odi.org.uk/resources/download/2844.pdf> (accessed 26 May 2009).

Jones, N, et al. (2008)

Improving Impact Evaluation Coordination and Use. A Scoping Study commissioned by the DFID Evaluation Department on behalf of NONIE. Available at: <http://www.odi.org.uk/resources/download/3177.pdf> (accessed 8 June 2009).

Kaiser, T (2004)

'Promise and Practice: Participatory Evaluation of Humanitarian Assistance' in *Forced Migration Review 8: 8–11*.

- Leeuw, F, and J Vaessen (2009)** *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. Draft Version for Discussion at the Cairo conference March–April 2009. Available online at: http://www.worldbank.org/ieg/nonie/docs/Guidance_IE.pdf (accessed 8 June 2009).
- Nelson, J (2008)** *Are we Ready for RCTs?* International Rescue Committee. Available at: <http://www.3ieimpact.org/doc/AreWeReadyforRCTs.pdf> (accessed 26 May 2009).
- NONIE Subgroup 2 (2008)** *NONIE Impact Evaluation Guidance*. Available at: http://www.worldbank.org/ieg/nonie/docs/NONIE_SG2.pdf (accessed 26 May 2009).
- NORAD (2008)** *The Challenge of Assessing Aid Impact: A Review of Norwegian Evaluation Practice Study 1/2008*. Oslo: Norwegian Agency for Development Cooperation.
- OECD-DAC (1999)** ‘Guidance for Evaluating Humanitarian Assistance in Complex Emergencies’, *Evaluation and Aid Effectiveness Series 1*. Available at: http://www.oecd.org/document/32/0,2340,en_2649_34435_1900640_1_1_1_1,00.html (accessed 8 June 2009).
- OECD-DAC (2002)** *Glossary of Key Terms in Evaluation and Results Based Management*. DAC Working Party on Aid Evaluation. Available at: <http://www.oecd.org/dataoecd/29/21/2754804.pdf> (accessed 8 June 2009).
- OIOS (2008)** *Review of Results-based Management at the United Nations*. Report of the Office for Internal Oversight Services (OIOS).
- Patton, M (1997)** *Utilisation-Focused Evaluation: The New Century Text*. Sage.
- Pawson, R (2003)** ‘Nothing as Practical as a Good Theory’ in *Evaluation* 9, pp. 471–490.
- Peberdy (2008)** *Impact Assessment Course Toolkit*. INTRAC.
- Proudlock, K, and B Ramalingam (2008)** *A New Agenda for News Media and Humanitarian Aid*. Report of the 23rd ALNAP Biannual Meeting 4 June 2008. London: ODI. Available at: http://www.alnap.org/meetings/pdfs/23_media.pdf (accessed 8 June 2009).
- Prowse, M (2007)** *Aid Effectiveness: The Role of Qualitative Research in Impact Evaluation*. ODI Background Note, London: ODI. Available at: <http://www.odi.org.uk/resources/download/430.pdf> (accessed 8 June 2009).
- Ramalingam, B, and J Mitchell with J Borton and K Smart (2009)** ‘Counting what counts: performance and effectiveness in the humanitarian sector’ in *ALNAP 8th Review of Humanitarian Action*. London: ODI.
- Robertson, D, et al. (2002)** ‘What Kind of Evidence do we Need to Justify Humanitarian Medical Aid?’ in *The Lancet*, Vol. 360, pp. 330–333.
- Roche, C (1999)** *Impact Assessment for Development Agencies*. Oxford: Oxfam GB.
- Roche, C (2000)** ‘Impact Assessment: Seeing the Wood and the Trees’ in *Development in Practice*, Vol. 10, No. 3 & 4, August.
- Roche, C (2008)** *The Seeming Simplicity of Measurement, Paper for the Ethical Questions for Non-Governmental Organisations Workshop*, 18–20 July. Melbourne: Oxfam Australia.
- Sandison, P (2006)** *The Utilisation of Evaluations in ALNAP Review of Humanitarian Action in 2005*. London: ODI. Available at: http://www.alnap.org/publications/RHA2005/rha05_Ch3.pdf (accessed 8 June 2009).
- Scriven, M (2008)** ‘A Summative Evaluation of RCT Methodology: An Alternative Approach to Causal Research’ in *Journal of MultiDisciplinary Evaluation*, Vol. 5, No. 9: 11–24.

- Sphere (2004)** *Humanitarian Charter and Minimum Standards in Disaster Response*. The Sphere Project: Geneva, Switzerland.
- Steckler, A, et al. (1992)** 'Toward Integrating Quantitative and Qualitative Methods: An Introduction' in *Health Education Quarterly* Vol. 19, No. 1, pp. 1–8.
- Stein, JG (2008)** "Humanitarian Organisations: Accountable – why, to Whom, for What, and How?" in Barnett, M and TG Weiss (eds) (2008) *Humanitarianism in Question: Politics, Power, Ethics*. Ithaca and London: Cornell University Press.
- Stern, E (2008)** *Current Thinking about Impact Assessment*. PowerPoint presentation delivered at the ALNAP 24th Biannual meeting. Available at: <http://www.alnap.org/meetings/24.htm> (accessed 8 June 2009).
- Watson, C (2008)** *Literature Review of Impact Measurement in the Humanitarian Assistance Sector*. Paper submitted to the Feinstein International Center, Tufts University for the Bill and Melinda Gates Foundation Project: Impact Assessment of Innovative Humanitarian Projects in Sub-Saharan Africa.
- Stansberry, M, and N Zagaria (2006a)** *DRAFT Concept Paper on Tsunami Recovery Impact Assessment & Monitoring System (TRIAMS)*. WHO/IFRC: Geneva, Switzerland.
- Stansberry, M, and N Zagaria (2006b)** *TRIAMS Regional Workshop Report*. United Nations, World Health Organization, and International Federation of Red Cross and Red Crescent Societies: Geneva, Switzerland. Available at: <http://www.ifrc.org/what/disasters/response/tsunamis/evaluation.asp> (accessed 26 May 2009).
- Stansberry, M, and N Zagaria (2007)** *Second TRIAMS Regional Workshop Report*. United Nations, World Health Organization, and International Federation of Red Cross and Red Crescent Societies: Geneva, Switzerland. Available at: <http://www.ifrc.org/what/disasters/response/tsunamis/evaluation.asp> (accessed 26 May 2009).
- White, H (2007)** *Evaluating Aid Impact*. MPRA Paper 6746 (posted 13 January 2008), Brighton: Institute of Development Studies, University of Sussex.
- White, H (2009)** *Some Reflections On Current Debates In Impact Evaluation*. 3ie Working Paper 1, 20 April. Available at: http://www.3ieimpact.org/admin/pdfs_papers/11.pdf (accessed 27 May 2009).
- Williams, K (1999)** *Mixing Qualitative and Quantitative Evaluation Tools: A Pragmatic Approach*. Centre for European Evaluation Expertise (C3E). Available at: <http://www.cemcentre.org/Documents/CEM%20Extra/EBE/EBE1999/Kevin%20Williams.pdf> (accessed 8 June 2009).
- WRC (Women's Refugee Council) (2009)** *Impact Evaluation in the Humanitarian Assistance Community*. Unpublished document.