EVALUATION OF HUMANITARIAN ACTION

PILOT GUIDE

ALNAP

ENTER ▶

EVALUATION OF HUMANITARIAN ACTION

About ALNAP

ALNAP is a unique sector-wide active learning membership network dedicated to improving the quality and accountability of humanitarian action.

The ALNAP membership works together to identify common approaches to improved performance; to explore new ways of improving learning and accountability through evaluations and other mechanisms; and to share ideas, lessons and innovations. Our overall aim is to make an active contribution to solving longstanding challenges facing the sector.

Our members are organisations and experts from across the humanitarian sector: donors, NGOs, the Red Cross/Crescent, the UN, independent consultants and academics.

www.alnap.org/EHA

USAID
FROM THE AMERICAN PEOPLE

ALNAP would like to acknowledge the financial support of USAID in carrying out this initiative.

Communications project management: Patrica Curmi
Designed by Deborah Durojaiye
Edited by Amanda A. Morgan

About the authors

Margie Buchanan-Smith is a Senior Research Associate with the Overseas Development Institute, and is based in London. She has been involved in the humanitarian aid sector for over 25 years, as an evaluator, a policy researcher and adviser, a humanitarian programme manager and as a trainer and coach. She is also currently a Visiting Fellow at the Feinstein International Center at Tufts University.

John Cosgrave is an independent evaluator based in Ireland. He has more than 30 years of experience of humanitarian action and development in over 60 countries. John was the evaluation advisor and coordinator for the Tsunami Evaluation Coalition and co-authored the synthesis report. He was also the lead author for the ALNAP pilot guide on Real-time evaluations of humanitarian action, published in 2009.

Margie and John co-facilitate annual training courses on 'Evaluation of Humanitarian Action' for Channel Research in Europe and for the World Bank-sponsored IPDET programme in North America.

FOREWORD

John Mitchell
ALNAP Director

As humanitarian evaluation practice becomes ever more vital in an increasingly diverse and complex system, this guide offers urgently needed support for evaluation specialists and non-specialists alike. Evaluation practice has evolved considerably, and much technical guidance already exists. Uniquely, however, this ALNAP pilot guide consolidates much of the current knowledge about every stage of a humanitarian evaluation: from initial decision to final dissemination.

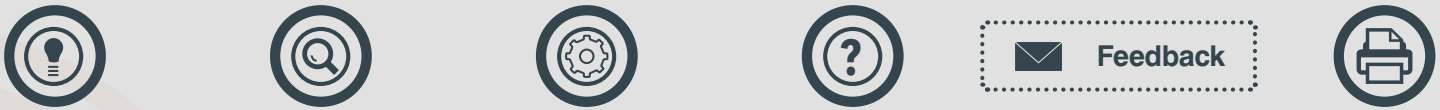
The guide format is a user-friendly and accessible interactive PDF. It contains real-life examples, practical tips, definitions and step-by-step advice on specific elements of evaluations at different stages of the process.

We want to ensure that the guide is as practical as possible for both evaluation specialists and the general reader and we are committed to testing it thoroughly. As always, we will harness the potential in the ALNAP network by inviting ALNAP Member agencies to help us pilot the Guide in their evaluative work. In this way, we know we can improve the quality, relevance and utility of the guide and ultimately help bring about better outcomes for people affected by crises.

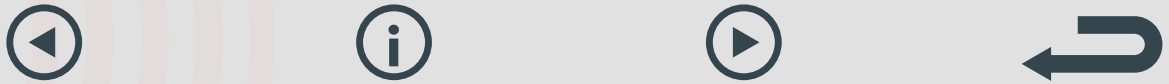
John Mitchell

HOW TO USE THIS INTERACTIVE GUIDE

This guide uses icons throughout. Rollover the icons below to find out what they mean.



At the bottom of each page in this PDF you will see a series of icons.



Move to previous page

Move to ‘How to use’ page

Move to next page

Return to previous page

This Interactive guide is best viewed with Adobe Reader. You can download the latest version free at http://get.adobe.com/uk/reader/

This guide can be viewed on tablets and smartphones, but some navigation features may be disabled.

Working offline – No internet connection is required to use the guide: we encourage you to download this interactive PDF and use in Adobe Reader. Feedback and viewing references do require an internet connection.

The EHA Guide is now also available in French and Spanish. These can be downloaded here: **www.alnap.org/eha**

SIX REASONS FOR THIS GUIDE

Here are six reasons we think it's time for a comprehensive EHA guide:

1. **Official donor assistance for humanitarian action has increased** *nearly six times in real terms from 1990 to 2011.*
2. **More interest and investment in evaluations** *as concerns are raised about effectiveness of development aid and humanitarian relief.*
3. **A critical mass of collective knowledge now exists to build on** – *ALNAP's evaluation database alone contains over 500 covering the last decade.*
4. **Commissioning of evaluations has shifted from agency headquarters to field-based staff** *as agencies decentralise - yet field-based managers often have little experience in planning and managing evaluations, especially EHA.*
5. **Little evidence that evaluation results lead to change of, or reflection on, policy and practice** – *better designed evaluations could provide more compelling evidence for policy change and promote utilisation.*
6. **The demand for guidance on EHA is growing** – *a Humanitarian Practice Network member survey in 2009 found that the number one guidance material requests were for EHA.*

This ALNAP guide provides practical and comprehensive guidance and good practice examples to those planning, designing, carrying out, and using evaluations of humanitarian action. The focus is on utilisation: to encourage you to consider how to ensure from the outset that an evaluation will be used.

This guide attempts to support high-quality evaluations that contribute to improved performance by providing the best evidence possible of what is working well, what is not, and why. The ultimate goal is to better meet the needs of people affected by humanitarian crises, who will be referred to throughout this guide as the affected population.

Affected population
All of the people affected (usually negatively) by a humanitarian crisis. This is a more neutral term than beneficiary, which implies that the target population received a benefit from the humanitarian intervention. Unfortunately, this is not always the case.

WHO IS THIS GUIDE FOR?

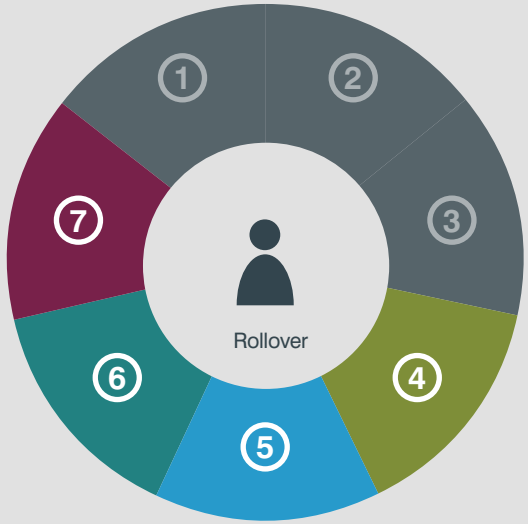
This guide has been written for all those directly involved in carrying out EHA.

Table of contents

Rollover the numbers and images below to find how a section applies to you.

Evaluation manager

Programme manager



Consultant

WHEN TO USE THIS GUIDE

Figure 1: The evaluation project circle

This chart shows how the guide follows the phases of the evaluation – planning, implementation, and dissemination of outputs.

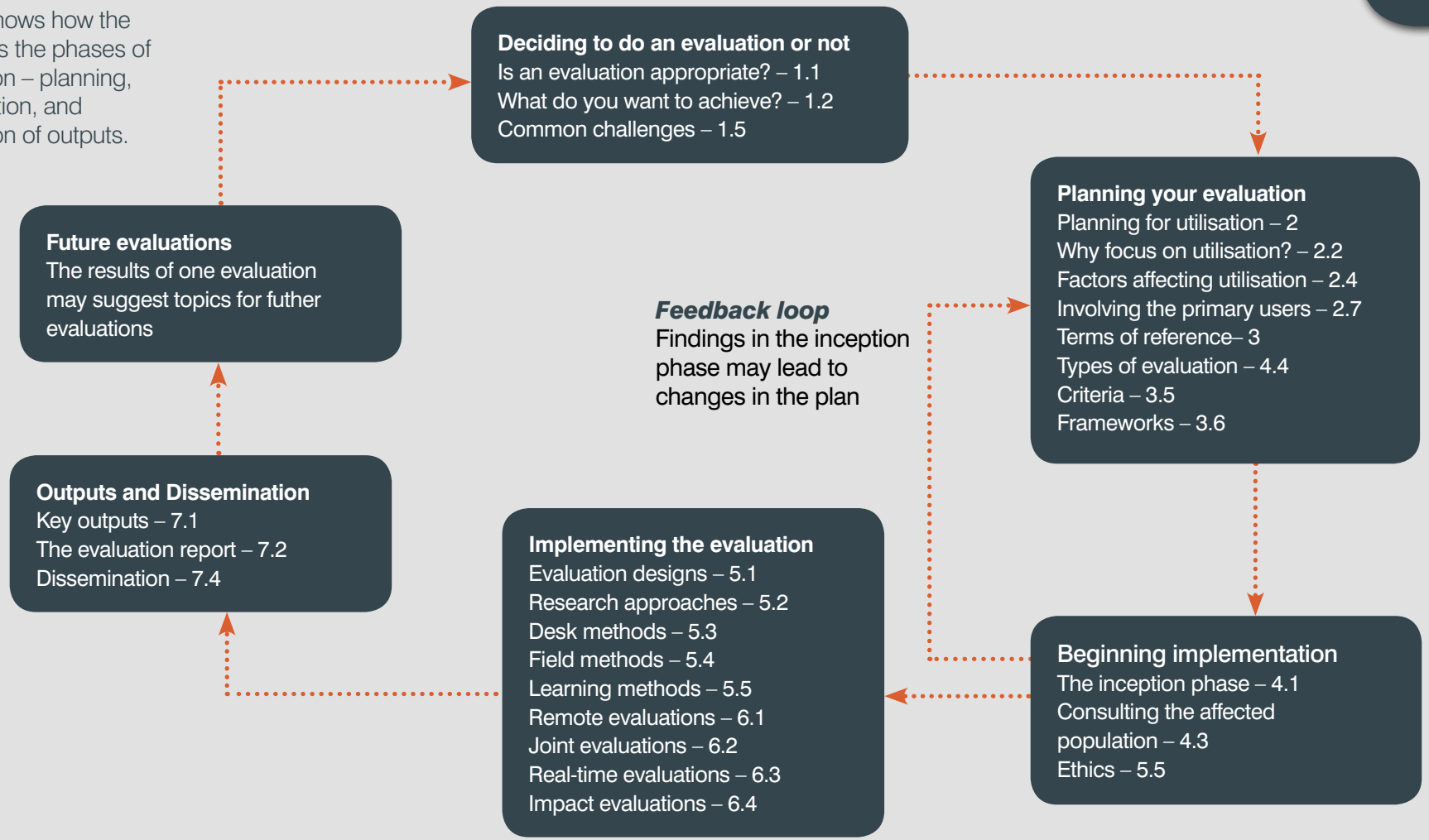


Table of contents

Feedback

PILOT AND ADDITIONAL PROTECTION COMPONENT

Table of contents

To further strengthen the utilisation- and user-focused dimensions of this guide on Evaluating Humanitarian Action (EHA) we are committing, over 2013 and into 2014, to **an active pilot process of this current draft version.**

We are working with a core group of pilot organisations from across the humanitarian system to use the guide in a range of EHA activities: from managing and conducting whole evaluation exercises – or single process steps part of a large EHA initiative – to facilitating EHA evaluation training and capacity-building activities.

If you are not one of our core piloters, we encourage you to let us know your experience of using the guide. No registration is required to participate, we simply ask that anyone who downloads this pilot version shares their comments on using any section, or part of a section, in their work. This can be done by clicking the confidential feedback buttons located at the bottom of guide pages.

If you would like to become more actively engaged as a core pilot organisation – which could involve online surveys, phone conversations and a roundtable discussion – the easiest way to get in touch with us, is via email at **eha@alnap.org**

www.alnap.org/EHA

During the pilot process, ALNAP will work to further populate the guide with more evaluator insights and good practice examples relating to the **evaluation of humanitarian protection outcomes**. The objective is to contribute to developing guidance to strengthen the evidential quality of protection-focused humanitarian evaluation. If you are interested in engaging with this initiative, or if you have some evaluation experience to share, specifically looking at how to apply a protection-lens to humanitarian evaluation, please contact us at **eha@alnap.org**

Acknowledgements:

The authors acknowledge the valuable feedback provided by members of an advisory group representing evaluation stakeholders in the humanitarian aid sector. Members of that group who have provided valuable feedback include Jock Baker, Tony Beck, Hana Crowe, Stefan Dahlgren, Wendy Fenton, Enrique Garcia, Josse Gillijns, Saul Guerrero, Babar Kabir, Peter Klansoe, Caroline Loftus, Rob McCouch, Joakim Molander, Jonathan Patrick, Nicoletta Pergolizzi, Riccardo Polastro, Hannah Reichardt, Peta Sandison, Vivien Walden.

The development of this guide was facilitated by the ALNAP Secretariat: John Mitchell, Paul Knox-Clarke, Francesca Bonino and, previously, Yuka Hasegawa. Communications support was provided by Patricia Curmi and Alexandra Warner.

TABLE OF CONTENTS

HOW TO USE THIS INTERACTIVE GUIDE 4

WHO IS THIS GUIDE FOR? 6

WHEN TO USE THIS GUIDE. 7

PILOT AND ADDITIONAL PROTECTION COMPONENT 8

1 DECIDING WHETHER TO DO AN EVALUATION 13

1.1 Humanitarian action 14

1.2 Evaluation of humanitarian action . . . 15

1.3 Weighing evaluation and other options 18

1.3.1 Accountability-oriented evaluations. 18

1.3.2 Learning-oriented evaluation 20

1.4 Evaluation in the response cycle. . . . 22

1.5 Common challenges 24

2 PROMOTING UTILISATION AND ENSURING QUALITY 31

2.1 Focusing on quality and utilisation at the outset. 32

2.2 What it means to be utilisation-focused 32

2.3 How evaluations of humanitarian action are used 33

2.4 Factors affecting utilisation. 35

2.5 Stakeholders 38

2.6 Focusing on what primary stakeholders need to know. 39

2.7 Involving the intended users of the evaluation. 40

2.8 Factors affecting evaluation quality . . 41

3 PLANNING AND DESIGN 44

3.1 Learning- and accountability-oriented evaluations 45

3.2 Types of evaluation 47

3.2.1 Evaluations with different scopes. 48

3.2.2 Evaluations of impacts or processes 49

3.2.3 Evaluations with different types of timing 50

3.2.4 Evaluations of and by different actors. 51

3.2.5 Theory-based evaluations 53

3.2.6 Other evaluation types 54

3.3 Determining knowledge needs 55

3.4 Evaluation criteria 57

3.5 Evaluation frameworks 62

3.6 Evaluation methods 69

3.6.1 Quantitative and qualitative approaches 69

3.6.2 Consulting the affected population 70

3.6.3 Document review 70

3.7 Management and governance. 71

3.7.1 Advisory groups 71

3.7.2 Managing conflict 73

3.8 Choosing an evaluation team 74
3.9 Identifying external consultants 78
3.10 Budgeting for an evaluation 81
3.11 Negotiating terms of reference 84

4 DOING AN EVALUATION OF HUMANITARIAN ACTION 92

4.1 The inception phase. 93
4.1.1 Preparing the inception report. . . 93
4.1.2 The evaluation matrix 97
4.1.3 Using a logic model 99
4.1.4 The desk review 104
4.1.5 Assessing the inception report . 104
4.2 Leadership and teamwork challenges. 108
4.3 Involving the affected population . . 111
4.4 Ethical issues. 113

5 EVALUATION DESIGN AND METHODS 119

5.1 Evaluation design. 120
5.1.1 Experimental designs 121
5.1.2 Statistical designs. 124
5.1.3 Theory-based designs 124
5.1.4 Case-based designs. 125
5.1.5 Participatory designs 125
5.1.6 One-off versus longitudinal designs 126
5.1.7 Choosing a design 126
5.2 Research approaches 127
5.2.1 Methods. 127
5.2.2 Sampling 129
5.2.3 Controlling research quality. . . . 133
5.2.4 Building an evidence base 138
5.2.5 Using triangulation 139
5.2.6 From data collection to analysis 140
5.2.7 Counterfactuals 142
5.3 Desk review methods. 142
5.3.1 Key document research 142
5.3.2 Using large document sets 144
5.3.3 Building a chronology. 146
5.3.4 Content and keyword analysis 149
5.3.5 Numerical analysis 152
5.3.6 Cost-effectiveness and efficiency 153
5.3.7 Assessing attention to gender in desk studies 155
5.3.8 Online surveys 157
5.4 Field methods 158
5.4.1 Key informant interviews 158
5.4.2 Interviews with the affected population 161
5.4.3 Observation 167
5.4.4 Participatory rapid appraisal methods 168
5.4.5 Field surveys 171
5.5 Learning-oriented approaches 176
5.5.1 After-action reviews. 177
5.5.2 Storytelling. 178
5.5.3 Appreciative inquiry 181

6 REMOTE, JOINT, REAL-TIME, AND IMPACT EVALUATIONS 185

6.1 Evaluation when access is constrained 186

6.2 Joint evaluations 188

6.2.1 Benefits of a joint evaluation . . . 188

6.2.2 How to plan and manage a joint evaluation. 189

6.2.3 What approach and methods to use. 190

6.2.4 How to ensure follow-up 191

6.3 Real-time evaluations 192

6.3.1 Advantages 192

6.3.2 Key characteristics 192

6.3.3 Methods. 196

6.4 Impact evaluations 198

6.4.1 Joint impact evaluations 199

6.4.2 How to do an impact evaluation 199

7 OUTPUTS, DISSEMINATION AND TAKE-UP 204

7.1 Key outputs 205

7.2 The evaluation report 209

7.2.1 Front matter 209

7.2.3 Recommendations 213

7.2.4 Annexes. 214

7.3 Circulating and commenting on the draft report 215

7.4 Approving and finalising the evaluation 217

7.5 Dissemination 218

7.6 Facilitating take-up of an evaluation 220

7.7 Evaluation syntheses, thematic reviews, and meta-analyses 221

ANNOTATED BIBLIOGRAPHY

DECIDING WHETHER TO DO AN EVALUATION

Contents of this section

1.1 Humanitarian action **14**

1.2 Evaluation of humanitarian action **15**

1.3 Weighing evaluation and other options **18**

1.4 Evaluation in the response cycle **22**

1.5 Common challenges **24**

How to use this section

This section is especially useful for evaluation and programme managers deciding whether or not to launch an evaluation. Sections **1.1** and **1.2** explain key concepts of humanitarian action evaluation in general, and EHA in particular. Section **1.3** reviews the types of goals for which an evaluation is (and is not) the best option. Section **1.4** explains how evaluation relates to monitoring and how it fits into the emergency response cycle, and Section **1.5** discusses how EHA differs from other types of evaluation and highlights common challenges and possible solutions.

Tables, figures and boxes

Box 1: Key evaluation concepts **16**

Table 1: Evaluation and other accountability processes **19**

Table 2: Evaluation and the three building blocks of a learning organisation **21**

Figure 2: Monitoring and evaluation in the emergency response cycle **23**

Figure 3: Sample management response matrix **28**

1.1 Humanitarian action

There are a number of definitions of humanitarian action. One of the most comprehensive and widely used was provided by the Good Humanitarian Donorship Initiative (2003):

> ***Humanitarian action***
> Action taken with the objective of saving lives, alleviating suffering, and maintaining human dignity during and after human-induced crises and natural disasters, as well as to prevent and prepare for them.

The parameters of what is called humanitarian action have gradually expanded. It used to be thought of simply as saving lives, but the importance of saving livelihoods is now widely accepted as well. Our working definition refers to maintaining human dignity, and the ability to support oneself is an important part of that.

Protection of civilians is now regarded as a responsibility of the whole humanitarian community and not just the Red Cross and the United Nations; it is about more than the provision of relief resources such as food, water and sanitation, shelter, and health services.

Humanitarian action includes not only response to a crisis but also support for preparedness and disaster risk reduction before a crisis and rehabilitation and recovery afterwards. The latter is a grey area that often falls between humanitarian and development aid. There is a growing realisation of the importance of addressing recovery needs immediately after a natural disaster. In conflicts and other protracted crises, it is often unclear when the emergency ends and recovery begins; in practice, both types of support are often needed and provided simultaneously.

Humanitarian action should be guided by the principles of humanity, impartiality, neutrality, and independence[1]. It is both *complicated* (with many different inter-related parts) and *complex* (non-linear and taking place in a dynamic environment with significant tensions and a requirement for judgment) (Glouberman 2002). This guide uses the term complexity to refer to both of these aspects.

[1] A Red Cross perspective on the humanitarian principles can be found in **Wortel (2009)**; **Leader (2000)**.

1.2 Evaluation of humanitarian action

EHA can be defined as 'the systematic and objective examination of humanitarian action, intended to draw lessons to improve policy and practice and enhance accountability'. A closer look at some of the key terms in this definition reveals the following:

- Systematic implies a planned and consistent approach.
- Objective implies stepping back from the immediacy of the humanitarian action and getting some perspective on it.
- Examination implies exploration or analysis to determine the worth or significance of the action.
- Drawing lessons to improve policy and practice and enhance accountability are the reasons for doing an evaluation, the use to which it will be put.

The extent to which an evaluation is truly independent depends on its purpose (see Section 4.2). This is most critical for accountability-oriented evaluations. It may be less achievable in learning-oriented evaluations if those doing the learning are involved in the evaluation and were responsible for implementing the humanitarian action being evaluated. However, even in such cases, some level of objectivity should be brought into the process, for example by having an external facilitator or experienced and independent resource people involved in or leading the team. Reducing bias should be a goal in all evaluations.

Two key types of evaluations, which are often contrasted, are evaluations for accountability and for learning.

Accountability
The process of taking responsibility for an intervention and accounting for it to different stakeholders, including those who are affected by the intervention, those who finance it, and other humanitarian actors.

Learning
The process through which experience and reflection lead to changes in behaviour or the acquisition of new abilities.

Box 1: Key evaluation concepts

Outputs
Defined deliverables provided by the actor or actors being evaluated.

Outputs are the result of inputs received and activities developed by the actor. An output must be fully attributable to an actor or group of actors – for example, water points provided in a camp of internally displaced people.

Outcomes
The uses made of the outputs by the affected population.

An outcome is only partly attributable to an actor – for example, the use made of the water from newly installed water points (for domestic consumption, animal consumption, or other livelihood activities such as brick making).

Impact
The social, economic, technical, or environmental effect of a project on individuals, gender and age groups, communities, and institutions.

Impacts can be positive or negative, macro (sector-level) or micro (household-level), intended or unintended. For example, providing food aid may prevent malnutrition (a positive, intended impact), but in some circumstances it may also discourage local food production (an unintended impact).

Impact in the results chain
The longer-term effects of the use of outputs by beneficiaries.

This term has a narrower definition than impact. For example, the impact of the freely available drinking water may be that it reduces illness in the internally displaced population or that it attracts other population groups to the camp. The actors who initiate an intervention contribute to the impact in the results chain, but it must also be attributed to a variety of factors outside the actors' control.

Attribution
The assertion that a particular input has caused a particular impact.

In complex and complicated humanitarian interventions, it is rarely possible to attribute a result to one specific cause. A food agency may attribute reduced malnutrition to food distribution, but it may also be due to improved water quality, child-care practices, hygiene, health care, sanitation, and vector control, or even normal seasonal changes.

Contribution
The extent to which a given input is thought to have helped to achieve a particular impact.

It is usually much easier in evaluation humanitarian action to assess contribution than attribution.

Consider the scope of your evaluation carefully. What do you want to focus on? Are you more concerned about the first phase of the emergency response or a later phase)? Do you want to look at preparedness as well? support to recovery and rehabilitation? The more focused your evaluation, the more usable your results are likely to be.

Accountability-oriented evaluation
An evaluation of how well resources have been applied (also called a summative evaluation).

Accountability and learning are difficult to separate in real life. Audits, which have a very strong accountability focus, may also lead to learning. Similarly, improved behaviour through learning is likely to improve accountability.

1.3 Weighing evaluation and other options

The primary focus of an evaluation, usually either accountability or learning, will help determine the choice of evaluation design (Section 4.2).

1.3.1 Accountability-oriented evaluations

Evaluation is one of several processes that can fulfil an organisation's responsibility to report to others – such as its board of directors, donors, or the affected population. You need to decide if it is the most appropriate and cost-effective option.

Table 1 summarises different types of accountability and suggests other options in addition to evaluation. The challenge in EHA is balancing accountability to a number of different stakeholders, with different interests and objectives and widely varying levels of power. For example, accountability to the affected population is seen as highly important, yet in many conflict environments, the affected population is disempowered and has a weak voice; in practice, donor organisations are among most powerful stakeholders.

Table 1: Evaluation and other accountability processes

Type of accountability	Is evaluation appropriate?	Other accountability processes
Strategic accountability (for example, to agency's mandate and objectives)	Yes	Strategic review
Managerial accountability (for example, for use of resources within an agency)	Yes	Performance management and other management tools
Financial accountability (for example, to donors)	Yes, especially if cost-effectiveness and efficiency are the main concerns	Audit (may be more appropriate if financial control and compliance are the main concerns)
Contractual accountability (for example, to carry out contracted tasks)	Yes, especially if there is a contractual obligation to do an evaluation	Audit; other processes specified in contract
Relational accountability (for example, to other agencies involved in an operation)	Yes, if this is included in the terms of reference for the evaluation	Institutional review
Legal accountability (for example, to local or international laws)	No	Legal (for example, labour law) compliance review
Accountability to the affected population	Yes, if the evaluation is designed appropriately	

Accountability often has both external and internal dimensions. The management of an organisation may be accountable for the use of resources not only to donors but also to the board of the organisation, as poor use of resources may threaten the survival of the organisation. Similarly, accountability to the affected population may reflect obligations that the management and the board have entered into (see **HAP, 2010**). If accountability is the main purpose of your evaluation, ask yourself the following:

- Which type of accountability are you principally concerned about?
- Is an evaluation the best way of fulfilling this?

1.3.2 Learning-oriented evaluation

Learning-oriented evaluation
An evaluation designed to facilitate individual, group, or organisational learning (also called *formative* evaluation).

Learning-oriented evaluations can be highly effective in examining what worked, what didn't, and how performance can be improved. But it is not the only available learning process, or the most cost-effective. A number of these are reviewed in Section **5.5**; they can be carried out by themselves or integrated into a learning-oriented evaluation.

An evaluation alone will not lead to learning by the organisation, although those who engage in and with the evaluation may learn as individuals. A *learning organisation* (**Table 2**) is most likely to benefit from an evaluation.

Table 2: Evaluation and the three building blocks of a learning organisation

Building block	Distinguishing characteristics	Implications for evaluation
A supportive learning environment	Staff members feel safe to disagree with others, ask naïve questions, own up to mistakes, and represent minority viewpoints; they recognise the value of opposing ideas, and are willing to take risks and explore the unknown.	Use evaluation methods that encourage openness—such as appreciative inquiry, the most-significant-change technique, and after-action reviews.
Concrete learning processes	Formal processes exist for generating, collecting, interpreting, and disseminating information.	Consider asking staff members to serve on the evaluation team. Establish a formal process for involving staff in analysing findings and drawing conclusions.
Leadership that reinforces learning	The organisation's leaders demonstrate their willingness to entertain alternative viewpoints; signal the importance of spending time on problem identification, knowledge transfer, and reflection; and engage in active questioning and listening.	Provide strong management support for evaluation. Provide adequate time for evaluation. Create space for feedback and the discussion of results with senior management.

Source: Garvin et al., (2008).

1.4 Evaluation in the response cycle

EHA is usually a one-off activity, undertaken at a key point in the humanitarian emergency response cycle in order to inform that cycle as well as future responses[2]. In some cases, a series of evaluations may be planned for different stages of the response cycle, as in the Darfur evaluation (**Broughton et al., 2006**). Monitoring, on the other hand, should be on-going throughout a humanitarian action. Monitoring and evaluation are complementary tools for helping determine how well an intervention is doing.

Monitoring
An on-going process of data collection; helps to answer questions of how much and how many; focuses on inputs and outputs.

Monitoring is generally conducted by those implementing a programme to ensure that it remains on track[3]. For example, monitoring of an emergency food aid programme would likely focus on inputs, such as the quantities of food aid delivered, and on outputs, such as the number of people receiving food aid. Occasionally it also captures outcomes – for example, rates of malnutrition and market prices of food. Monitoring is often weak in humanitarian action, and the lack of good monitoring data often creates problems for evaluation.

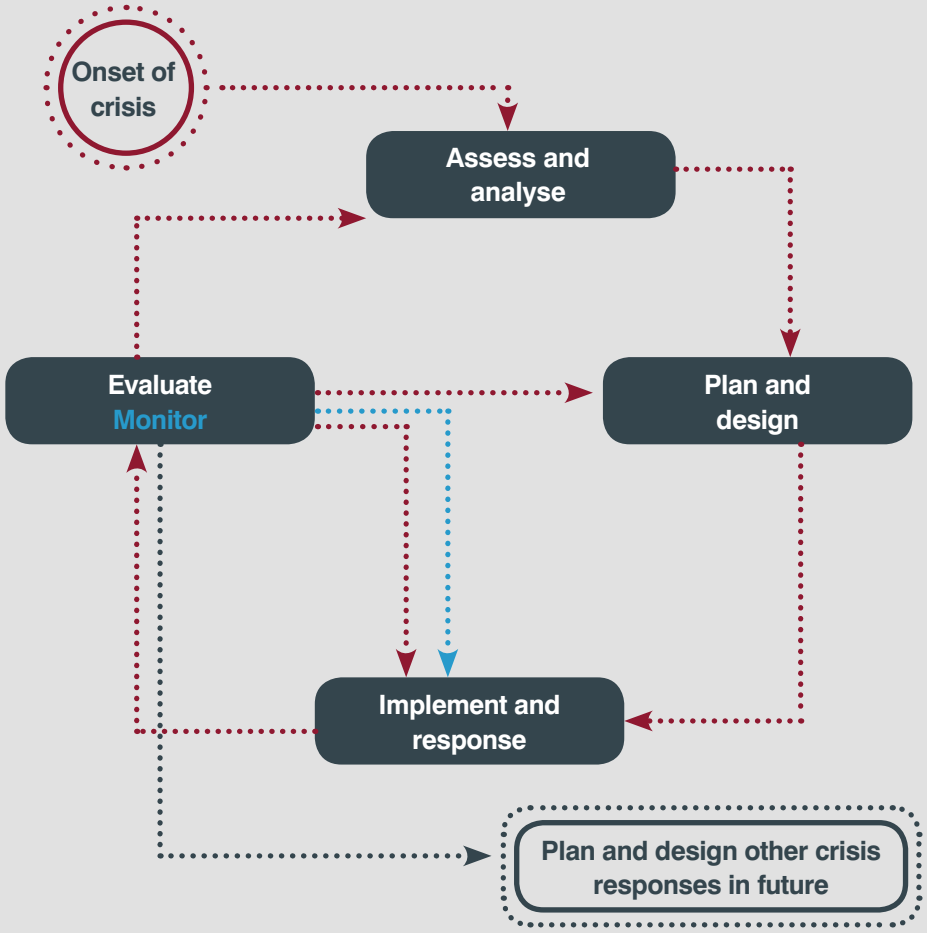
Evaluation
The episodic assessment of the performance of a programme or project; helps to answer questions of what and why; focuses on outcomes and impact.

An example of impact would be the effect of an emergency food aid programme on local food markets. Evaluation focuses on what results were achieved or not achieved with the inputs and outputs at hand – on how well organisations are meeting their overarching objectives and on what might need to change for them to do so. It may be carried out by internal or external staff, but is often conducted by people not involved in programme delivery, in order to give it more credibility with stakeholders.

[2] This section is based on **Imas Morra and Rist (2009)** and **Hallam (2011)**.

[3] External forms of monitoring, such as Citizen Report Cards (**World Bank, 2004**), exist but are generally found in development rather than humanitarian contexts.

Figure 2: Monitoring and evaluation in the emergency response cycle



To ensure that an evaluation is useful and is used, it should be scheduled when its results can best contribute to key decision-making moments within the response cycle (**Figure 2**).

It is not appropriate to commission evaluations for every project or programme if this is likely to overload the organisation and if effective monitoring systems are in place. Strategic evaluations are more likely to fulfil key knowledge needs in a timely manner.

Strategically selected evaluations are fundamentally different from the cyclic approach described in **Figure 2**.

Key questions to consider are:

- how can the evaluation process add value to the organisation as a whole?
- how many strategic evaluations does the organisation have the capacity to absorb?

The evaluation unit of the Swedish International Development Cooperation Agency has successfully experimented with this approach.

Good practice example 1

Strategic selection of evaluations

The Swedish International Development Cooperation Agency's evaluation planning cycle starts with the evaluation unit having conversations with all operational units, to determine what they would like to know and how evaluation could help. A list of around 100 evaluation ideas is initially generated, from which the evaluation department chooses 15. These are in addition to about 80 evaluations carried out each year at the operational level (**SIDA, 2011**).

It may be particularly appropriate to conduct an evaluation:

- For a programme with unknown or disputed outcomes
- For large and expensive interventions
- For pilot initiatives
- Where the agency has a strategic interest
- Where stakeholders are keen on an evaluation

It is inappropriate to conduct an evaluation:

- When it is unlikely to add new knowledge
- When security issues or lack of data would undermine its credibility

1.5 Common challenges

The same crisis circumstances that necessitate humanitarian action also make EHA different from other forms of evaluation, particularly evaluation of development work, and often more challenging[4]. Evaluators should make clear the constraints they face from the outset of the evaluation, and how they intend to deal with them, from a risk management perspective. The final report should indicate how these constraints affected the evaluation process and findings. The following list describes common EHA challenges and their potential solutions.

- **The urgency and chaos of humanitarian emergencies** – A rapid humanitarian response is, by definition, planned quickly and often in extremis. Planning and monitoring documents may be scarce, objectives may be unclear, and early plans may quickly become out-dated as the context changes or is clarified.

  Potential solution: Consider constructing a chronology of the crisis and of the humanitarian action to be evaluated. Use interviews with agency staff to identify actual or implicit objectives and how these might have changed over time.

[4] Material in this section was adapted from **Beck (2006)**.

- **Lack of baseline data** – Data may have been destroyed in the disaster; they may also have become irrelevant – for example, when a large proportion of the population has been displaced.

  Potential solution: Conduct interviews to ask beneficiaries and local key informants about the extent to which conditions have changed and the reasons for the changes. Include similar questions in surveys. See Section 5.4 and **Box 7**.

- **Staff turnover** – The high turnover of staff working in humanitarian action, especially international staff, can make it difficult for evaluators to find and interview key informants.

  Potential solution: Invite staff and ex-staff key informants who are absent to participate in a telephone interview or online survey (see Section 5.3.8).

- **Polarised perspectives** – Conflicts often intensify differences in perspective; events are subject to widely differing interpretations, making 'objective' evaluation difficult.

  Potential solution: Gather as many different points of view as possible. Ensure that different views are represented in the final evaluation report, although the evaluation team may be required to give judgement on aspects of the humanitarian response based on the information gathered. See Section 5.2.5.

- **Breakdown in trust due to politicisation or trauma** – In conflicts and other humanitarian crises that become politicised, and where there is widespread abuse, trauma, and fear, trust breaks down within the affected population, and it may be hard to get in-depth and accurate responses to questions.

  Potential solution: Design data collection methods and ways of engaging with the affected population that are sensitive to trauma and fear. Consider gender-sensitive interviewer-interviewee pairing. See Sections 4.3 and 4.4.

- **Insecurity and lack of access** – Security issues may make it difficult or impossible, especially in conflict environments, for evaluators to reach the affected population.

  Potential solution: Explore creative ways of carrying out the evaluation remotely. See Section 6.1.

- **Lack of staff time to talk to evaluators** – Humanitarian organisations and their staff may be struggling to implement programmes in highly stressful environments. Making time to spend with evaluators may be a low priority. The operational culture of humanitarian action can also be a disincentive to reflection.

  Potential solution: Some short reflective-learning exercises can yield good insights if well facilitated. See Section 5.5.1. Ask staff what they want to know about their operation and in which areas they are keen to change how the organisation approaches a task.

- **Lack of clear lines of responsibility** – The lack of clearly defined responsibilities among humanitarian actors can hinder accountability and attribution of impact, especially in a large-scale crisis involving many humanitarian actors.

  Potential solution: Focus on contribution rather than attribution. Consider carrying out a joint evaluation to explore the combined impact of a number of agencies, or of the entire international humanitarian response, rather than attempting to artificially isolate the impact of any one agency. See Section 5.1.

- **Unrealistic workloads** – The terms of reference may set out a two-person-year task to be done with a budget for four person-months.

  Potential solution: Use an inception report to refine the evaluation task so that it can be completed in the allotted period. This is a useful way of managing expectations early on. See Section 4.1.1.

Inception
Initial phase of an evaluation, during which the evaluation team tries to develop a full understanding of the evaluation task and prepares a detailed report outlining the plan for the evaluation.

- **Remote location and damaged infrastructure** – These challenges can make access difficult.

  Potential solution: Carefully plan the fieldwork with someone who is familiar with the current situation, including likely travel times and access constraints, or consider letting the partner plan the fieldwork visits subject to criteria set by the evaluation team.

- **Time pressure on the affected population** – People may have little time to participate in an evaluation because of their own urgent survival tasks.

  Potential solution: Avoid time-consuming methods such as focus groups, and use group evaluations instead. Consult with beneficiaries while they are waiting for supplies or transport. Use direct observation of beneficiary behaviour of affected populations. See Section 5.4.3.

- **Ethical constraints on experimental approaches** – The hypothesis "if the drought-affected people had not received food aid, many of them would have died" is often called a counterfactual, as (assuming that they did get food) it runs counter to the facts. Ethical considerations would, of course, prohibit testing this hypothesis by withholding assistance from some people.

  Potential solution: Compare different means of assistance (for example, cash grants, vouchers, and direct food assistance). Take advantage of any natural experiments that may have arisen, such as a group that got no assistance due to its isolation.

- **Lack of learning in the humanitarian community** – ALNAP reviews of humanitarian action find the same issues arising year after year, despite being identified in evaluations.

  Potential solution: Build in a focus on utilisation from the start of an evaluation. Consult the likely users and use methods that are credible to them. See Section 2. Encourage the development of a management response matrix (**Figure 3**) with regular follow-up by senior managers.

Management response matrix
A record of management's response to each evaluation recommendation and the steps managers plan to take to address it, with a target date and responsible party for each step.

Figure 3: Sample management response matrix

	Further funding required?		Management response			Comment	Action to be taken	Timing	Responsible unit
	Yes	No	Accept	Partially accept	Reject				
Recommendation 1									
Recommendation 2									

Source: FAO.

Some evaluation challenges can be addressed by carrying out an evaluability assessment beforehand to determine if the evaluation is feasible and how it would best be conducted.

Evaluability assessment
A review of the extent to which an activity or a program can be evaluated reliably and credibly.

It is much easier to organise a management response when recommendations are targeted at particular managers or levels of management.

Can this section be improved?

NOTES

NOTES

Please share your feedback:

Help us make the Guide more practical and user-oriented by sharing your feedback on its content and navigation. Feedback can include questions, suggestions, useful resources or practical advice from your own experience. You can share your comments and ideas by clicking on the feedback button at the bottom of any page. This will generate a confidential email to the EHA team in the ALNAP Secretariat. For more information on how you, your team or organisation can engage more actively in piloting the Guide contact us directly at **eha@alnap.org**

PROMOTING UTILISATION AND ENSURING QUALITY

Contents of this section

2.1 Focusing on quality and utilisation at the outset 32

2.2 What it means to be utilisation-focused 32

2.3 How evaluations of humanitarian action are used 33

2.4 Factors affecting utilisation 35

2.5 Stakeholders 38

2.6 Focusing on what primary stakeholders need to know 39

2.7 Involving the intended users of the evaluation 40

2.8 Factors affecting evaluation quality 41

2

How to use this section

This section explains what it means to be utilisation-focused and different ways in which an evaluation might be used (Sections 2.1 to 2.3). Sections 2.4 to 2.8 suggest ways to promote utilisation by identifying the primary stakeholders, involving them in the evaluation process, and taking steps to give the evaluation the best chance of being used. These are important considerations for managers commissioning and managing evaluations, and must be taken into account at the beginning of the evaluation process to succeed[5].

[5] This section is based on Patton (2008), Sandison (2007), and Hallam (2011).

Tables, figures and boxes

Table 3: Different ways to use evaluations of humanitarian action 34

Figure 4: Factors affecting utilisation of an evaluation of humanitarian action 36

Box 2: Stakeholder mapping 38

2.1 Focusing on quality and utilisation at the outset

Although many evaluations of humanitarian action have successfully assessed the merit and significance of different programmes and projects, and have identified key lessons for learning, their record in bringing about change is much weaker. They have tended to be somewhat disconnected from the culture and mind-set of humanitarian organisations. This threatens to undermine their credibility. Much greater attention must be given to how evaluations are used if they are to fulfil their potential:

A serious intention to use an evaluation is indicated by careful planning, adequate time and appropriate mechanisms for follow-up that suit and serve the users. *(Sandison, 2007)*

The quality of the entire evaluation process, not just the final report, is a critical factor determining its usefulness and the likelihood of it being used (see Section 2.8). Other factors are also essential to ensure utilisation (see Section 2.4), but quality is the factor over which the evaluation team has the greatest control.

2.2 What it means to be utilisation-focused

A utilisation-focused evaluation is done with the intended primary users in mind, for specific, declared, practical uses. When planning an evaluation, it is essential to identify the intended users at the start and to help them decide what they want to achieve with it. If this has not been done by the time an evaluation team is appointed, which is often the case, then the team needs to do it. To be most effective, this should involve repeated interactions between the users and the team.
User goals should guide the choice of approach and methods, which should also take into account constraints such as limits to time and resources. The utilisation focus and collaboration with primary users should continue to guide the evaluation process from planning through implementation. This way of thinking and working requires commitment and time.

Good practice example 2

Helping key stakeholders clarify their objectives

In 2006, a learning-oriented evaluation was carried out for the German church organisation Diakonie (**Welch and Otto, 2006**). Evaluators fully involved Diakonie's implementing partners in the field. They encouraged Diakonie to involve its local partners in preparing an inception report to identify evaluation objectives and methods. This shared learning approach was fed into the field research and the joint insights gained then shared back up the line. Participatory workshops were held at the initiating (joint orientation) and concluding (shared interpretation) stages of each phase.

2.3 How evaluations of humanitarian action are used

EHAs are used in two ways: (1) for accountability (also called summative) purposes, and (2) for learning, usually focused on a specific project (formative), but sometimes focused on general, system-wide knowledge (developmental). Typically an evaluation contains a mix of elements, though one may predominate (**Table 3**).

Table 3: Different ways to use evaluations of humanitarian action

Use	Questions	Examples
Summative – judging the merit or worth of a programme (for example, to fulfil its accountability to stakeholders or inform funding decisions)	Does the programme meet needs? Does it have merit? What are its outcomes?	The British Department for International Development's *Multilateral Aid Review (2011)* examined how well the work of different agencies matched donors' priorities.
Formative – to enhance learning (for example, to improve a programme)	What works and what doesn't? What are current strengths and weaknesses?	*The Organisational Learning Review of Caritas Internationalis' Response to the Tsunami Emergency* (**Otto et al., 2006**) facilitated the process of learning, emphasising openness and the participation of key stakeholders.
Developmental – to enhance learning (for example, introducing new ideas to an organisation or to the sector as a whole)	Does the programme take real world events and limitations into account? What are general patterns across programmes?	The Joint Rwanda Evaluation (**Borton et al., 1996**) introduced new ideas about the accountability of humanitarian agencies and precipitated major policy innovations such as the Sphere standards.

Note: based on **Patton (2008)**.

Contents
1
2
3
4
5
6
7
Feedback

Sometimes evaluations are not used effectively. This can take several forms:

- **Ritualistic use** – the evaluation serves a purely symbolic purpose, for example to fulfil a contractual obligation to the funder but with little or no commitment on the part of the agency to using it.
- **Misuse** – findings are suppressed, misrepresented, or distorted to serve a personal or political agenda.
- **Non-use** – findings are ignored because users find little or no value in them, or they are not aware of them, or the context has changed dramatically. Non-use may be a consequence of a lack of management buy-in, poor evaluation design, or an evaluation's failure to answer its own questions or to provide compelling evidence for its conclusions.

2.4 Factors affecting utilisation

Factors that affect utilisation fall into four broad categories:

1. **Quality** – for example, the quality of the inception process, data gathering, analysis, or final product. This is the most important factor affecting credibility.
2. **Organisational context** – for example, the organisational culture of learning and knowledge management, and the organisational structure in terms of the proximity of evaluation units to decision-makers. Evaluations are much more likely to be used in an organisation that actively seeks information on performance in order to improve programme management and delivery. This can be greatly influenced by whether key leaders support learning or place a low value on evidence-based decision-making and are inclined to be defensive in the face of evaluation.

3. **Relationships** – for example, between evaluators and evaluation users. The personal factor – for example, the presence of an evaluation champion(s) among the key stakeholders with whom it is critical for the evaluators to build and maintain a relationship – can be crucial in determining whether an evaluation is used or not. The formal position and authority of such champions may matter less than their enthusiasm and interest.
4. **External influences** – for example, the public or media, or indirect stakeholders whose actions can affect the use of an evaluation.

These are represented as a three-circle model in **Figure 4**, with external circumstances represented by the blue shaded area around the circles.

Figure 4: Factors affecting utilisation of an evaluation of humanitarian action

This model is based on the RAPID framework for assessing research and policy links. See http://www.odi.org.uk/rapid/tools/Documents/Framework.pdf.

Contents
1
2
3
4
5
6
7
Feedback

Good practice example 3

Elements of a highly influential evaluation

The Joint Evaluation of Emergency Assistance to Rwanda in 1996 (**Borton et al., 1996**) is one of the most influential evaluations of humanitarian action ever undertaken. Its success can be explained in terms of the three-circles model (**Figure 4**):

- This was a high quality evaluation, because of the thoroughness and rigour of the work, the calibre of the team, and the way it made recommendations. It had high credibility with potential users.
- In terms of organisational culture and structure, the international humanitarian system was open to change at that moment, partly because of widespread unease about the highly variable performance amongst humanitarian agencies. A number of well-placed individuals championed use of the evaluation and skilfully created or exploited alliances or networks on its behalf.
- In terms of relationships, the evaluators had strong ties to key policy-makers in the sector.
- External influences were also important. After the shock of the Rwanda crisis, and the intense media coverage, policy-makers were more open to change.

When planning a utilisation-focused evaluation, it is worth asking the following questions:

- How can we ensure that the evaluation asks questions whose answers are likely to be acted upon?
- How can we ensure that the evaluation is of a sufficiently high quality to have credibility with users?
- How is the organisational context likely to foster or hinder take-up of the evaluation's findings? What can we do about it?
- How can we develop the relational links, for example between evaluators and users, to ensure that the evaluation findings and recommendations are taken up?
- What external influences could support or hinder take-up of the evaluation's findings?

Ideally, each of the three circles should be strong and overlap with the others, and the external context should be conducive to the evaluation. If it is not possible to create these conditions, you may want to consider whether this is the best time to conduct an evaluation or whether an evaluation is the best tool for your purposes.

2.5 Stakeholders

Identifying the key stakeholders is a critical step in EHA planning and should be done early. Stakeholders are people who will be impacted by the evaluation findings. Some may have a direct interest – for example, funders, implementing staff, and the affected population. Others may have an indirect interest – for example, staff working for other humanitarian agencies, government officials, and taxpayers.

Stakeholder mapping is a useful way of categorizing the interests and needs of different groups of stakeholders. If done visually, this can also be a participatory process involving a number of people. See **Box 2**.

Box 2: Stakeholder mapping

Different groups of stakeholders can be mapped on a series of concentric circles, distinguishing between the primary stakeholders, who are expected to be fully engaged in the evaluation, and those with an indirect interest, who may play a different role. Those with an indirect interest are further divided between those who should be influenced by the evaluation as well as consulted and those who should merely be consulted.

One way of doing this type of mapping is to write down all the different stakeholders on sticky notes, draw circles on a flip-chart, and then place each note in the appropriate circle according to the stakeholder's interests and proposed engagement in the evaluation.

Helpful questions to ask about stakeholders include the following:

- Who are the stakeholders with a direct interest? Which of them are also the primary intended users of the evaluation?
- How can you engage the primary stakeholders, encourage their ownership of the evaluation, and ensure its relevance to their needs?
- Which of the primary stakeholders should receive the highest priority in order to keep the evaluation practical and focused?

The last question is particularly important if there are a number of primary stakeholders with different and competing interests – for example, funders who want to know about the effectiveness and cost-efficiency of the programme for accountability purposes, and programme staff who want to learn about what worked and what didn't and how they could do better. Unless you clearly identify and prioritise the primary stakeholders and intended users, the evaluation's competing purposes may become unmanageable, and you may not adequately achieve any of them.

2.6 Focusing on what primary stakeholders need to know

Once you have identified the primary stakeholders and intended users, ask them: What do you need to know to enable you to better decide what to do and how to do it?

This question encourages a utilisation focus and minimises the risk of receiving an unfocused shopping list of questions. There may be uncomfortable issues that programme managers need to know about in order to improve programming. Ask stakeholders what they will do with the information they expect to get from the evaluation.

Map primary stakeholders and group them by presumed areas of common interest to simplify the process of asking them what they will do with evaluation findings.

Once you know the answer to the core question, follow up with these questions:

- Who needs what and why?
- What is the purpose of this evaluation?
- What exactly is being evaluated?

2.7 Involving the intended users of the evaluation

People are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings. And they are more likely to understand and feel ownership if they've been actively involved throughout. Ways of accomplishing this include the following:

- Find out what people want to know and why when first designing the evaluation (Section 3.3).
- Involve potential users in compiling the terms of reference (Section 3.11).
- Consider having some users join the evaluation team (Section 3.9).
- Form a reference group for the evaluation that includes key users (Section 3.7.1).
- Hold a workshop in which users and the evaluation team design the evaluation together.
- Ensure that the evaluation team regularly communicates with users throughout the evaluation process and that the users are involved in key decisions, for example about refocusing the evaluation or making trade-offs.
- Hold a workshop to present evaluation findings. Consider asking the users to participate in drafting recommendations.
- Ask the users to design a dissemination strategy (Section 7.5).

Evaluations can provoke anxiety or resistance amongst those whose work is being evaluated. This may be exacerbated in high-profile humanitarian crises that have attracted international media attention. In these cases, building a sense of ownership is even more important. Frequent communication throughout the evaluation can also help. Building a sense of ownership among the evaluated may have to be balanced with the need to preserve objectivity. Engendering a commitment to evaluation often involves promoting openness to change.

2.8 Factors affecting evaluation quality

The quality of an evaluation depends on six key factors: design, participation, planning, the end products, follow-up mechanisms, and evaluator credibility[6]. If you can answer 'yes' to all the questions below, you are on track for conducting and producing a high quality evaluation.

- **Design** – Is the purpose of the evaluation clear, and has it been agreed among key stakeholders? Has the approach been designed according to the purpose and users' needs and interests? (For example, for a learning-oriented evaluation, are field staff involved?) Will the design and methods provide sufficient evidence to answer the evaluation questions? Are appropriate mechanisms in place to ensure that threats to the independence of the evaluation are dealt with?
- **Participation and ownership** – Will key stakeholders be involved throughout the evaluation?
- **Planning** – Is the lead time sufficient for staff to be adequately involved in the evaluation? Is the timing of the evaluation appropriate to influence decision-making?
- **Products** – Are the reports and other products accessible and easy to read and understand? Is the evidence credible (well-researched, objective, expert)? Are the methods clearly described, and are any inherent limitations acknowledged? Is there a clear logical flow from evidence to findings, conclusions, and recommendations? Are recommendations clear, specific, prioritised, constructive, measurable, relevant, time-bound, and targeted?
- **Follow-up mechanisms** – Were specific follow-up plans established at the outset?
- **Evaluator credibility** – Are the evaluators credible in terms of competence and reputation? Does the evaluation team have the necessary mix of skills and experience for this evaluation? Are the evaluators sufficiently independent (and able to maintain that independence during the evaluation)? Do they have appropriate interpersonal skills? Are they able to remain balanced, impartial, objective, and constructive?

Can this section be improved?

[6] This section is based in part on Sandison (2007).

NOTES

NOTES

Please share your feedback:

Help us make the Guide more practical and user-oriented by sharing your feedback on its content and navigation. Feedback can include questions, suggestions, useful resources or practical advice from your own experience. You can share your comments and ideas by clicking on the feedback button at the bottom of any page. This will generate a confidential email to the EHA team in the ALNAP Secretariat. For more information on how you, your team or organisation can engage more actively in piloting the Guide contact us directly at **eha@alnap.org**

PLANNING AND DESIGN

Contents of this section

3.1 Learning- and accountability-oriented evaluations 45
3.2 Types of evaluation 47
3.3 Determining knowledge needs 55
3.4 Evaluation criteria 57
3.5 Evaluation frameworks 62
3.6 Evaluation methods 69
3.7 Management and governance 71
3.8 Choosing an evaluation team 74
3.9 Identifying external consultants 78
3.10 Budgeting for an evaluation 81
3.11 Negotiating terms of reference 84

3

How to use this section

Section 3.1 provides advice on identifying the purpose of an evaluation, while Section 3.2 focuses on choosing the best type of evaluation. Sections 3.3 and 3.4 identify knowledge goals and evaluation criteria, while Section 3.6 helps to decide how specific to be in identifying an evaluation framework and methods. Sections 3.7 to 3.10 explore the managing, contracting, and budgeting for an evaluation, and Section 3.11 focuses on establishing terms of reference.

Tables, figures and boxes

Table 4: Accountability and learning-oriented evaluations 46
Box 3: Top level questions 56
Table 5: Evaluation frameworks 64
Figure 5: UNICEF's conceptual framework for malnutrition 67
Box 4: Evaluating the human resource component of humanitarian action 68
Table 6: Advantages and disadvantages of using internal and external evaluators 75
Table 7: Advantages and disadvantages of using individuals or a contractor's team 79
Table 8: Cost elements for an evaluation 83
Table 9: Elements of the terms of reference 85

3.1 Learning- and accountability-oriented evaluations

As described in Section **1.3**, there are two principal reasons for doing an evaluation: accountability and learning. Most practical evaluations address both to some extent, but one should predominate. A common problems with EHAs is not choosing a primary focus and giving both equal weight. It is important to be clear about the primary focus, as this can influence the evaluation style and choice of methods.

For example, an accountability-oriented evaluation is likely to adopt a more adversarial investigative style, seeking to attribute responsibility for both success and failure. But this may not be conducive to learning if it inspires defensiveness on the part of those who need to learn. Learning needs an environment of psychological safety (Edmondson, 2004) where it is permissible to acknowledge difficulties and admit mistakes.

Participatory and facilitated evaluations are a more appropriate approach when learning is the primary goal, because those who want or need to learn are more centrally involved in the thinking and reflective process. **Table 4** shows how the two approaches might differ.

Table 4: Accountability and learning-oriented evaluations

Evaluation element	Accountability-oriented evaluation	Learning-oriented evaluation
Terms of reference	Based on input from external stakeholders as well as programme management	Should be set by those directly involved in the programme who want or need to learn
Team membership	Independent external staff	Internal staff, perhaps with an external facilitator or leader, or mixed internal and external staff
Emphasis in approach	Methodology of data collection and analysis (more objective)	Process of reflection and reaching conclusions (more participatory)
Methods	Mix of quantitative and qualitative methods that will provide robust evidence	Participatory methods involving those who are to learn
Management	Those responsible for accountability	Those responsible for knowledge management and learning
Management style	More directive	More facilitative
Circulation of report	In the public domain	May be limited to the organisation itself to encourage open and honest participation

Contents

When planning and designing an evaluation, these are the key questions to ask:

- Which purpose is more important – accountability or learning? Where on the continuum between accountability and learning would you place the purpose? It should not be in the middle, or you risk losing focus.
- If accountability is the main purpose, accountability to whom and for what?
- If learning is the main purpose, learning by whom?

When you have clarified the main purpose, you can design an evaluation process and methods to achieve it.

3.2 Types of evaluation

Once you have decided who the primary stakeholders and intended users are, what they want and need to know, and the overall purpose of the evaluation, then you must decide on the type of evaluation. Considerations include the following:

- The scope of the evaluation – for example, whether it is focused at the project, programme, or sector level
- Whether to focus on processes or outcomes
- Whether to emphasise accountability or learning
- The timing of the evaluation – real-time, mid-term, or after the end of the intervention
- How many actors will be evaluated
- Whether the evaluation will be based on programme theory (see Section 3.2.5)
- Any other distinguishing aspects of the evaluation

Evaluations typically belong to several categories simultaneously. The evaluation type, or combination of types, also helps determine the evaluation design (see Section 5.1).

3.2.1 *Evaluations with different scopes*

Most evaluations are focused at the project, programme, cluster, or sector level.

Project evaluation
Evaluation of a single humanitarian intervention with specific objectives, resources, and implementation schedule, which often exists within the framework of a broader programme.

One example of this type of evaluation is the evaluation of Tearfund and Tear Netherland's shelter projects after the 2009 earthquake in Padang, Sumatra (Goyder, 2010).

Programme evaluation
Evaluation of a set of interventions with a unifying humanitarian objective.

One example of this type is the evaluation of Danida's assistance to internally displaced people in Angola (Cosgrave, 2004). Programmes sometimes emerge from a set of interventions with coherent objectives, as in this example, rather than being designed from the start as a coherent programme, as was the evaluation of UNICEF's education programme in Timor-Leste (Tolani-Brown et al., 2010).

Cluster evaluation
Evaluation of multiple projects within a larger programme.

Examples of a *cluster evaluation* include AusAid's evaluation of a cluster of NGOs after the Pakistan earthquake (Crawford et al., 2006) and of NGO work in the Pacific (Crawford and Eagles, 2008).

The term *cluster evaluation* also sometimes refers to the evaluation of the UN Cluster Coordination System, such as the initial evaluation by Stoddard et al., 2007, an evaluation approach developed by the Kellogg Foundation in the 1980s in which the evaluator accompanies the disparate projects in a programme beginning when the grant is awarded and promotes learning in those projects (Barley and Jenness, 1993; Sanders, 1998; WK Kellogg Foundation, 1998, p. 17).

Sector evaluation
Evaluation of a group of interventions in a sector, all of which contribute to the achievement of a specific humanitarian goal. The evaluation can cover part of a country, one country, or multiple countries.

One example of a sector evaluation is an evaluation of the water, sanitation, and hygiene sector in the Democratic Republic of the Congo (van der Wijk et al., 2010).

3.2.2 *Evaluations of impacts or processes*

Impact evaluation
An evaluation that focuses on longer-term effects rather than aid delivery (usually, but not always, carried out some time after project completion).

One good example of an impact evaluation is the evaluation of the US State Department's Bureau of Population, Refugees and Migration's assistance to Burundi refugees (Telyukov et al., 2009).

Another excellent example is the impact assessment of community-driven reconstruction in Lofa County (Fearon et al., 2008). All evaluations that look at the results of an intervention are forms of impact evaluation.

Process evaluation
An evaluation that focuses on the processes by which inputs are converted into outputs; may also examine the intervention as a whole.

Process evaluations have the advantage that they can take place long before results are clearly evident; many real-time evaluations contain a strong element of process evaluation. The joint evaluation of the humanitarian response to the Rwanda crisis (Borton et al., 1996) was primarily a process evaluation, even though it also looked at results. Most humanitarian evaluations are a mixture of process and impact evaluation. One special type of process evaluation is the normative evaluation.

Normative evaluation
An evaluation that compares what is being implemented with what was planned or with specific standards.

Normative evaluations are a subset of process evaluations. They may evaluate project implementation against the project plans; in EHA they are more likely to evaluate against standards. This was the case with the DEC evaluation of the Gujarat earthquake response, which used the NGO Code of Conduct (SCHR and ICRC, 1994) as a benchmark (Humanitarian Initiatives UK et al., 2001).

3.2.3 Evaluations with different types of timing

Evaluations fall on a continuum that runs from the almost completely retrospective (backward looking) to the prospective (forward looking). The timing of the evaluation determines how strong these two elements can be.

Evaluability assessments are useful when it is not clear if a credible evaluation will be possible – for example, the Somalia evaluability assessment (Cosgrave, 2010). They are sometimes combined with initial reviews of programmes, as in the Programme Review and Evaluability Study of UNICEF's Education in Emergencies and Post-crisis Transition Programme (Barakat et al., 2010) and the preparatory review for the evaluation of the Global Education Cluster (Reid et al., 2010).

Real-time evaluation
An evaluation of an on-going humanitarian operation as it unfolds.

Examples of real-time evaluations include the evaluations of the Pakistan Swat Valley crisis (Cosgrave et al., 2010) and the 2010 Pakistan floods (Polastro et al., 2011b).

Mid-term evaluation
An evaluation performed towards the middle of an intervention.

One example is the mid-term evaluation of DG ECHO's Regional Drought Decision for the Greater Horn of Africa (Wilding et al., 2009).

Ex-post evaluation
An evaluation performed after an intervention has been completed.

One example is CARE's evaluation of its tsunami relief response in two districts of Sri Lanka (Bhattacharjee et al., 2007).

On-going evaluation
A series of evaluations designed to run throughout an intervention.

One example of an on-going evaluation is the pair of IASC evaluations of the response to the Haiti earthquake at 3 and 20 months after the earthquake (Grünewald et al., 2010; Hidalgo and Théodate, 2012). Another example is a UNHCR (United Nations High Commissioner for Refugees) evaluation that ran for 6 years after the end of an intervention (Skran, 2012).

Ex-ante evaluation
An evaluation performed before an intervention begins.

Such evaluations, rare in the humanitarian sector, are based on the lessons learned from previous such operations. One example is the Review Concerning the Establishment of a European Voluntary Humanitarian Aid Corps (Bruaene et al., 2010).

3.2.4 *Evaluations of and by different actors*

Self-evaluation
An evaluation by those who design and deliver an intervention.

Most such evaluations are internal documents; one published example is the evaluation of Medair's response to the 2004 Asian tsunami in Sri Lanka, which was conducted by a member of the initial team in Sri Lanka who later returned to carry out the evaluation (Lee, 2005).

Participatory evaluation
An evaluation in which stakeholders, including beneficiaries, work together to design, carry out, and interpret an evaluation.

Participatory evaluations are relatively rare in the humanitarian sector; one example is a Colombia country study (Mendoza and Thomas, 2009) for the UNHCR evaluation of age, gender, and diversity mainstreaming (Thomas and Beck, 2010).

Single-agency evaluation
An evaluation carried out by the agency that implemented the intervention.

Examples of this common type of humanitarian evaluation include CARE's evaluation of its Pakistan earthquake response (Kirkby et al., 2006) and Oxfam's evaluation of cash transfers in Sudan (Bush and Ati, 2007).

Joint evaluation
An evaluation carried out by two or more agencies, evaluating the work of two or more agencies.

This might involve a donor and recipient agency, multiple agencies with similar missions, or different actors working in the same sector. Examples include joint evaluations carried out by NGO partners in an emergency capacity-building project after the Yogyakarta earthquake (Wilson et al., 2007).

System-wide evaluation
An evaluation of the international humanitarian system's response to a humanitarian crisis, open to all actors in the system.

One example of this is the Tsunami Evaluation Coalition's evaluation of the international response to the Indian Ocean tsunami of 2004 (Telford et al., 2006). Another example is the IASC evaluation of humanitarian assistance to South Central Somalia (Polastro et al., 2011a).

Theory-based evaluation
An evaluation based on the underlying programme theory.

Programme theory
A theory or model of how an intervention contributes to specific outcomes through a series of intermediate results.

3.2.5 Theory-based evaluations

Theory-based evaluation has mostly been used in evaluations in the development sector, especially for education and public health projects. Explicit theory-based evaluations are rare in the humanitarian sector. Boller et al. (2011) take a theory-based approach in their evaluation of UNICEF's early childhood development support. Programme theory is sometimes referred to by other names, including programme logic (Funnell and Rogers, 2011). Identifying the programme theory is critical for theory-based evaluation.

Logical framework
A matrix that identifies inputs, outputs, outcomes, and impacts, their causal relationships, the underlying assumptions, indicators, and the levels expected to be achieved.

Realist evaluation
A form of theory-based evaluation that pays close attention to the context of the intervention and regards the target of the assistance not as a monolithic group but as a collection of distinct sub-groups.

The evaluation uses the logical framework for the programme as the framework for the evaluation. See Section 4.1.3 for more detail on theory-based evaluation and logical frameworks.

Realist evaluations do not ask "what works?" or "does this programme work?" but "what works for whom in what circumstances and in what respects, and how?" (Pawson and Tilley, 2004). Thus, a realist evaluation would not examine how different subgroups within an affected population are served by an intervention. For example, the *Report on the Links between Relief, Rehabilitation and Development in the Tsunami Response* (Christoplos, 2006) noted that only groups with fishery-based livelihoods got assistance to restore their livelihoods.

3.2.6 Other evaluation types

Funding mechanism evaluation
An evaluation of a set of interventions funded by a particular mechanism.

One example of such an evaluation is the UN Food and Agriculture Organisation's evaluation of its projects funded by the Central Emergency Response Fund (CERF) (Cossée et al., 2010).

Technology evaluation
An evaluation of a specific technique or technology.

These evaluations are not common; one example is the review of the use of the British Red Cross Mass Sanitation Module after the 2010 Haiti earthquake (Fortune and Rasal, 2010). Such evaluations can also cover innovative approaches such as the use of Community-based Management of Acute Malnutrition in Nepal (Guerrero, 2010).

Institutional evaluation
Evaluation of the internal dynamics of implementing organizations; their policy instruments, service delivery mechanisms, and management practices; and the linkages among these.

Examples of this type of evaluation are the two evaluations of UNICEF's programme for humanitarian capacity-building, funded by the UK Department for International Development (DFID) (Bhattacharjee et al., 2010; Brown et al., 2005), and the Independent Review of UNICEF's Operational Response to the January 2010 Earthquake in Haiti (Bhattacharjee et al., 2011).

Policy evaluation
An evaluation that examines the understandings, beliefs, and assumptions that make individual projects possible as well as desirable; may evaluate both the efficacy of the policy itself and how that policy has been implemented.

These are often quite wide-ranging and tend to be reviews rather than evaluations. Examples are the UN's *Humanitarian Response Review* (Adinolfi et al., 2005), DFID's *Humanitarian Emergency Response Review* (Ashdown, 2011), and the UN Office for the Coordination of Humanitarian Affairs (OCHA's) *To Stay and Deliver: Good Practice for Humanitarians in Complex Security Environments* (Egeland et al., 2011).

Meta-evaluation
Either an evaluation designed to aggregate findings from multiple evaluations, or an evaluation of the quality of one or more evaluations.

One example is ALNAP's meta-evaluation of joint evaluations (Beck and Buchanan-Smith, 2008). Scriven (2011) provides a checklist for meta-evaluations of the second type.

Thematic evaluation
An evaluation of a selection of interventions that all address a specific humanitarian priority that cuts across countries, regions, and possibly agencies and sectors.

One example of a thematic evaluation is the evaluation of the role of needs assessment after the December 2004 Asian earthquake and tsunamis (de Ville de Goyet and Morinière, 2006). This was one of five thematic studies for the Tsunami Evaluation Coalition. The recent UNHCR review of refugee education is an example of a single-sector thematic evaluation (Dryden-Peterson, 2011).

3.3 Determining knowledge needs

The expense of an evaluation is only justifiable if it provides needed, usable information – not just what you would like to know, but what you need to know to improve programming, information that will make a difference to what you do (see Section 2.6).

There are two levels of questions in an evaluation: top-level questions that appear in the terms of reference (ToR) and that the evaluators are tasked to answer, and the questions that the evaluators then ask of others to gather the data needed to answer the top-level questions. The International Committee of the Red Cross limits evaluations to three top-level questions. This ensures that the most crucial questions get adequate attention.

The Red Cross used to request all interested parties to identify the range of questions and issues they would like included in an evaluation. The evaluation department then reframed this into 'evaluable' questions. However, it was found that the scope of the evaluation always grew, until it became difficult to manage the process. To mitigate this, the evaluation department now tries to focus on just three key questions for each evaluation. (Hallam, 2011, p. 11)

One of the biggest problems with many current ToRs in EHA is they spell out questions in too much detail and delve too much into the second level of questions.

Box 3: Top level questions

Top-level questions should focus on the information needed for major policy and operational decisions. They may include the following:

- Did the intervention achieve its purpose, and if not, why not?
- Should we continue with this approach, and if so, why?
- What changes do we need to make to our system to be better able to respond to a similar crisis in the future?

Competent evaluators should be able, in the inception phase, to identify the key second-level questions that they need to ask in order to answer the top-level questions.

When you circulate a draft ToR, do not ask for suggested questions. This will invariably give you dozens of questions that will be very difficult to rationalise as every question is someone's favourite. Instead, ask stakeholders what information they want and what they would use that information for. Many such information requests can be incorporated into broader top-level questions.

Contents
1
2
3
4
5
6
7

3.4 Evaluation criteria

We engage in evaluation every day. Whether buying apples in the supermarket or picking a sandwich in the canteen, we make a judgement about the options before us based on the criteria of quality and value. A formal EHA uses a much more structured approach but essentially looks at similar criteria, although the term cost-effectiveness (comparing different approaches to the same task) is often preferred to value (which could imply a notional monetary value for human life and suffering).

These two fundamental criteria need to be broken down into a set of criteria that can be systematically assessed. In 1991, the Development Assistance Committee of the Organisation for Economic Cooperation and Development (OECD/DAC), focusing on the most common problems noted in development projects, proposed four quality criteria – relevance, effectiveness, sustainability, and impact – and the value criterion of efficiency (OECD/DAC, 1991). A few years later, the OECD/DAC had these criteria adapted for EHA in complex emergencies (OECD/DAC, 1999), adding the criteria of coverage and coherence, suggesting appropriateness as an alternative to relevance and connectedness as an alternative to sustainability, and proposing two new criteria: coordination and protection. These criteria reflected the biggest problems seen in humanitarian action in the 1990s.

Although the OECD/DAC criteria are not perfect, there are good reasons for using them instead of devising new criteria for each evaluation.

- Using the standard criteria makes meta-evaluation (the drawing of lessons from a wide range of evaluations) much easier.
- The standard criteria are likely to capture common weaknesses in humanitarian action, based on experience and research.
- Evaluations that use standard criteria are easier for evaluation managers and other evaluators to work with.

Choosing the best evaluation criteria is often the greatest challenge for new evaluators. Is a particular question one of effectiveness or relevance? Often the answer is a matter of personal preference or convention. There are various valid ways to break down the fundamental criteria of quality and value, and the subcategories often overlap.

When planning an evaluation, first identify what you need to know, then place it within the evaluation criteria – not vice versa. The criteria are tools to think with and may suggest additional relevant questions. Only ask evaluation questions if you are ready to take action based on the answers. Only use the criteria that relate to the questions you want answered. What matters are the questions, not the criteria.

On the next page we summarise basic evaluation criteria in the form of a real-word question, followed by a criterion and its definition, followed by an action that could be taken based on an answer to the question. These criteria take both positive and negative and intended and unintended impacts into account.

CRITERIA FOR EVALUATION OF HUMANITARIAN ACTION

Question: To what extent was the food package adapted to local needs?

Criterion: appropriateness

Appropriateness
How well humanitarian activities are tailored to local needs.

Appropriateness increases ownership, accountability, and cost-effectiveness. It is a variant on the traditional OECD/DAC criterion of relevance, the extent to which a project is in line with local needs and priorities and donor policy.

Action steps: Management may change planning procedures; operations may change the design of future aid packages.

Question: To what extent did the food distribution reduce malnutrition?

Criterion: effectiveness

Effectiveness
How well an activity has achieved its purpose, or can be expected to do so on the basis of existing outputs.

Action steps: Management may refine its policy for such interventions; operations may redesign the current programme and design future programmes differently.

Question: Was using partners for distribution more cost-effective than using contractors would have been?

Criterion: efficiency

Efficiency
A measure of the outputs, qualitative and quantitative, achieved as a result of inputs.

Evaluating efficiency usually requires comparing alternative approaches to achieving an output, to see whether the most efficient approach has been used. Action steps: Management may change or retain its policy on giving priority to partners; operations may negotiate fairer rates with partners.

CRITERIA FOR EVALUATION OF HUMANITARIAN ACTION

Question: What impact did the decision to provide food to the senior female in the household rather than the male head of household have on household dynamics?

Criterion: impact

Impact was defined earlier as any effect on an individual, group, community, or institution. It can be positive or negative, intended or unintended.

Action steps: Management may change its policy on mode of distribution; operations may fine-tune the distribution practice, for example to better take gender dynamics into account.

Question: How did the distribution of food assistance affect the level of planting for the next harvest?

Criterion: connectedness

Connectedness
The extent to which short-term emergency response steps take longer-term and interconnected problems into account.

Connectedness is related to the concept of sustainability, the idea that interventions should support longer-term goals and eventually be managed without donor input.

Action steps: Operations may improve the design of future programmes.

Question: To what extent did the food distribution target the most vulnerable?

Criterion: coverage

Coverage
The extent to which assistance reaches all major population groups affected by the crisis.

Action steps: Operations may improve the design of future programmes.

CRITERIA FOR EVALUATION OF HUMANITARIAN ACTION

Question: How did the decision to import sugar rather than purchase it from the national mill fit with the donor's policy on encouraging local purchases?

Criterion: coherence

Coherence
The extent to which there is consistency across security, developmental, trade, military, and humanitarian policies, and to which all policies take into account humanitarian and human-rights considerations.

Coherence is a policy-level issue that may not be relevant in single-agency, single-project evaluations.

Action steps: Management may seek better ways to balance cost-effectiveness with adherence to broad policies.

Question: Did gaps or duplication exist in food distribution by this agency and other agencies?

Criterion: coordination

Coordination
The extent to which different actors' interventions are harmonised, promote synergy, and avoid gaps, duplication, and resource conflicts.

Coordination is sometimes included in the effectiveness criterion rather than treated as a separate criterion.

Action steps: Management and operations may seek better ways to engage with other actors.

Beck (2006) provides additional details on the use of the standard OECD/DAC evaluation criteria, including discussion of the definitions and use of the criteria and examples of questions that address them.

3.5 Evaluation frameworks

Evaluations are challenging, and there is a danger that some vital aspect may be overlooked, that the evaluation may lack conceptual rigour, or that may not meet accepted standards. Frameworks can help with these issues in the following ways:

- They provide a structure for breaking up the evaluation task into smaller elements that are easier to deal with.
- They reduce the risk that key elements will be overlooked, by systematically directing the team's attention to all elements of the evaluation subject.
- They provide a structure for the evaluation that will be recognised by stakeholders. Stakeholders may have had little exposure to the standard evaluation criteria, but they are often familiar with the structure of conceptual frameworks or standards in common use in their sector.
- They can provide a baseline of generally accepted good practice against which the project can be tested.

The OECD/DAC criteria described in Section 3.4 offer one possible framework for an evaluation, but using them as a framework removes the focus on the evaluation questions.

Other frameworks may add specificity and detail to the evaluation and make it more accessible to users. Such frameworks can be used in addition to the OECD/DAC criteria, or may even be substituted for them if appropriate. Evaluators are often asked to use one of the following types of frameworks:

- Broad normative frameworks reflect the norms that define the humanitarian endeavour. They include international humanitarian law, the humanitarian principles, and various conventions.
- Conceptual frameworks are broad models of how a system is thought to work. They are often well tested and validated. Sources of conceptual frameworks include recent academic research, major evaluations, and project design documents.

- Standards and guides can be used both as standards to evaluate against and as a way to break down humanitarian actions into small components that are easier to examine. They include sector-specific guides and agency operations manuals and emergency manuals.

Frameworks and their uses are summarised in **Table 5**.

Table 5: Evaluation frameworks

Frameworks and examples	Possible use
Broad normative frameworks • International humanitarian law such as the Protocol on the Protection of Victims of Non-international Armed Conflicts (ICRC, 1977) • Humanitarian principles (Wortel, 2009) • Convention on the Rights of the Child (UN General Assembly, 1989)	Normative frameworks can be used during the inception phase to see if there are any aspects of the intervention that might raise concerns and need to be more closely examined. Some frameworks also provide checklists or standards against which to review policies and performance.
Broad conceptual frameworks • Utstein peace-building categories (Smith, 2004) • Livelihoods Framework for Complex Humanitarian Emergencies (Lautze and Raven-Roberts, 2003) • UNICEF Conceptual Framework for Malnutrition (Black et al., 2008)	Conceptual frameworks can be used to evaluate the success of each element of an intervention and the extent to which all elements work together. They can help give an evaluation intellectual coherence and ensure that all components are adequately addressed.
System-wide standards and guidelines • Principles and Good Practice of Humanitarian Donorship (Good Humanitarian Donorship, 2003). • NGO/Red Cross/Red Crescent code of conduct (Borton, 1994). • People in Aid Code (People in Aid, 2003). • Humanitarian Accountability Partnership standards (HAP, 2010). • Sphere standards (Sphere Project, 2011) • Guiding Principles on Internal Displacement (OCHA, 2004) • Fragile States Guidelines (OECD/DAC, 2007)	These standards can provide a checklist or reference point against which to evaluate performance, a basis for breaking down the evaluation into manageable tasks, and a structure for the report. They are most effective when an agency has made a formal commitment to adhere to them; otherwise, it may be difficult to justify using them. The inception phase is the best time to propose using a particular standard.

Frameworks and examples	Possible use
Sectoral standards • Sector-specific elements of the Sphere standards (Sphere Project, 2011) • Minimum Standards for Education in Emergencies (INEE, 2006)	Sectoral standards form a good basis for organising sectoral evaluations. In some cases, they are based on general standards; for example, the Sphere standard on consultation with affected populations is meant to apply to all sectors.
Agency standards and guides • United Nations High Commissioner for Refugees' Handbook for Emergencies (UNHCR, 2007) • World Food Programme's Emergency Field Operations Pocketbook (WFP, 2002) • UNICEF's Emergency Field Handbook (UNICEF, 2005)	These documents can provide a good basis for checking compliance (accountability) and also for breaking down and organising the evaluation task.

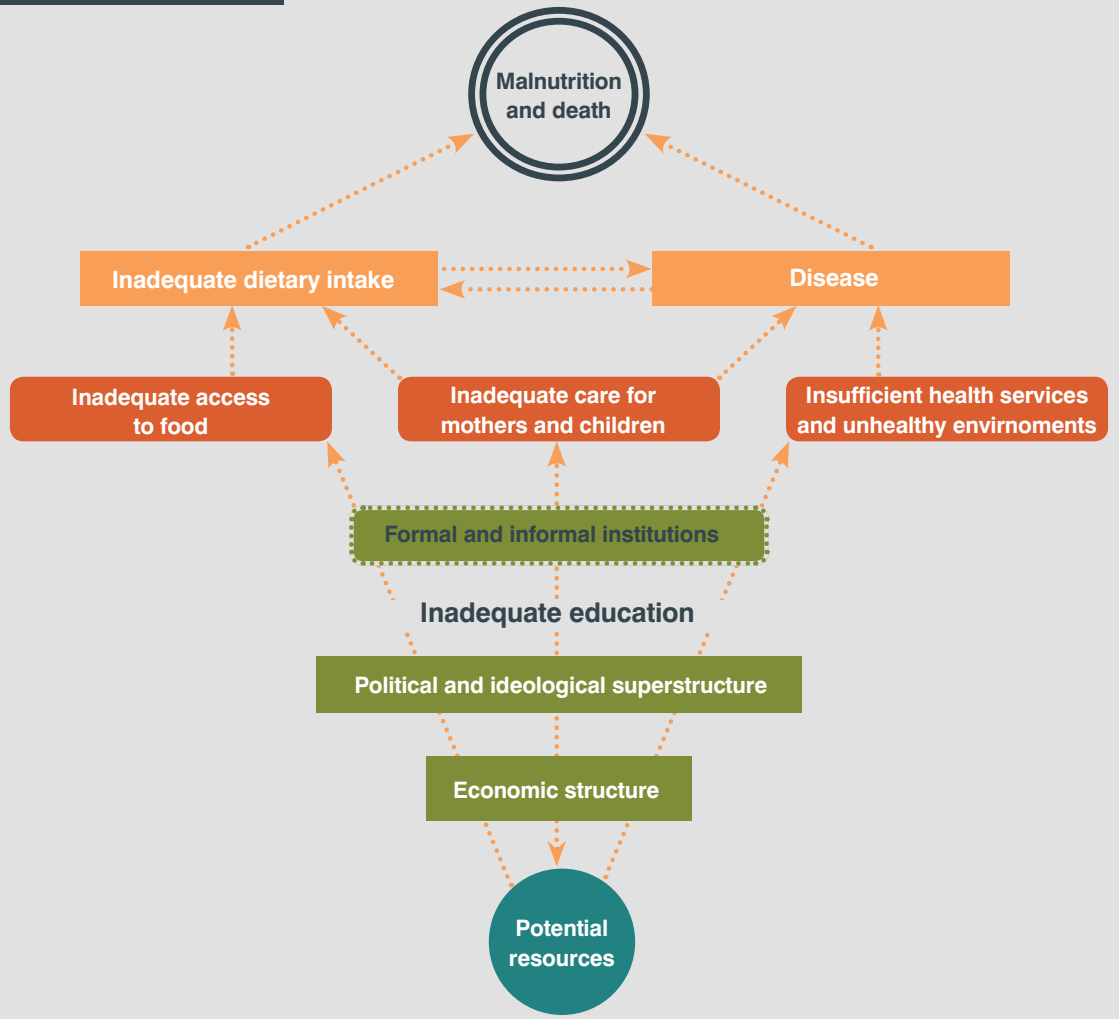
Ideally, the ToR should identify the conceptual frameworks and standards that the evaluators are expected to use. Alternatively, the evaluation team may want to propose a conceptual framework and relevant standards in the inception phase.

Key questions to consider include:

- Do project planning documents make any reference to particular conceptual frameworks or standards?
- Would it be appropriate to use any of these as a framework for the evaluation?

If the evaluation uses one framework that is central to the evaluation task, this, together with the evaluation criteria, can provide a good way to structure the report. However, if the evaluation uses multiple standards, it may be advisable to pick one as the main framework and address the others in short annexes to the main report.

Figure 5: UNICEF's conceptual framework for malnutrition



Contents
1
2
3
4
5
6
7

Feedback

An example of a conceptual framework is the UNICEF malnutrition framework presented in **Figure 5**. (Conceptual models are often, but not always, presented as diagrams.) The diagram shows three levels of causes: basic, underlying, and immediate. If you were evaluating a project intended to reduce malnutrition and related deaths, you might examine the extent to which the project explored and addressed at least the immediate and underlying causes. One example of the use of this framework was the SCF research into the causes of malnutrition in children under three in Ethiopia (Meyer, 2007).

Box 4 presents an example of the practical use of a framework, in which key second-level evaluation questions are based on the framework provided by the People in Aid Code of Good Practice (People in Aid, 2003).

Box 4: Evaluating the human resource component of humanitarian action

The staff of a humanitarian agency, both nationally and internationally recruited, are a key determinant of the success or failure of a humanitarian intervention, yet the human resource component of humanitarian action is rarely given much attention in EHA except when there are obvious problems. Key evaluation questions on human resource management include the following:

- To what extent is the staffing structure and capacity appropriate and adequate for effective implementation of operations?
- To what extent is management capacity appropriate and adequate for effective implementation of operations?
- To what degree have human resource systems, practices, and policies supported or hindered operations?
- How has leadership supported or hindered overall performance?

This box was written with input and guidance from People In Aid.

3.6 Evaluation methods

The selection of detailed evaluation methods (see Section 5) is normally the job of the evaluation team. However, the evaluation manager needs to consider some methodological issues, because they affect the time and budget needed for the evaluation.

It is appropriate for an evaluation manager to specify that the evaluation team is expected to use both quantitative and qualitative methods in a rigorous way. There may also be times when the evaluation manager needs to specify that a particular method should be used – for example, for institutional reasons or compatibility with other evaluations.

Normally, within whatever limits are set by the evaluation manager, the evaluation team's inception report will propose the specific methods to be used, pending the evaluation manager's approval. In large evaluations, a peer review group often also comments on the proposed methodology.

An evaluation manager who specifies a particular method accepts responsibility for the results if that method is unsuitable for answering the evaluation question in that particular context. It is generally better to specify either the methods or the results that you want, not both.

3.6.1 Quantitative and qualitative approaches

Evaluation methods are divided into two broad categories: quantitative and qualitative. Sometimes evaluation managers are pressured to use one or the other.

Quantative method
A method used for collecting numerical data.

If a quantitative study is well designed, its findings can be generalised for a larger population using statistical methods (see Section 5.2). Because of their origin in the natural sciences, quantitative methods are sometimes seen as more scientific than qualitative methods. Quantitative methods predominate in the natural sciences, the oldest branch of science. However, their basic premises – that observations are independent of the observer,

and that facts are separate from theories – are problematic in social research. Qualitative methods do not rely on such premises and benefit from a strong literature addressing such interactions.

Qualitative method
A method used for collecting non-numerical data, typically descriptions of meaning and experience.

Most EHAs rely on mixed methods with qualitative methods predominating, because such methods are better able to answer typical evaluation questions and because of the technical difficulties that the rigorous use of quantitative methods poses in the complexity of a typical humanitarian context (see Section **5.4.5**).

3.6.2 Consulting the affected population

There are a number of options for consulting the affected population (see Section **5.4**). For example, you could conduct a quantitative survey, or you could use a detailed qualitative survey carried out by a social researcher. You may wish to conduct interviews or use participatory rapid appraisal techniques.

Detailed performance monitoring data may be available for some projects. Are you going to consult the affected population directly or through the administrative structures? Are you just consulting, or do they have a direct stake in the evaluation and are engaged from the outset?

Allow extra time for the fieldwork if you want the evaluation team to conduct detailed consultations with the community rather than the quick visits that are the norm. Consultation with the affected population has long been a weakness in EHA (see Section **4.3**).

3.6.3 Document review

Considerations when planning a document review include (1) what are the relevant documents, and how can you make them available? and (2) how much time should be allocated to the document review? This is a key part of the evaluation that is often not given the time it requires. Section **5.3** covers desk study methods, with attention to specific tasks such as content analysis and working with large document sets.

3.7 Management and governance

The evaluation manager's task begins when the decision is made to conduct or consider conducting an evaluation. In some cases, an agency's evaluation policy or plan may make this decision routine; in others, the evaluation manager may need to lobby stakeholders to gain support for an evaluation.

A field visit by the evaluation manager prior to the evaluation can be useful for fine-tuning the ToR and for getting support for the evaluation from stakeholders in the field.

3.7.1 Advisory groups

The best arrangements for managing an evaluation ensure that primary stakeholders remain engaged in its decisions. Advisory groups are often formed for this purpose. The most common types are discussed below.

STEERING GROUP

Steering group
A group established to steer an evaluation through key stages such as establishing the ToR, writing the inception report, and drafting the final report.

A steering group typically consists of evaluation managers drawn from a number of agencies. Strong leadership is essential to enable the group to work effectively and to its full potential, especially if membership is diverse.

MANAGEMENT GROUP

Management group
A group that manages the evaluation on a day-to-day basis, including drafting the ToR, contracting and managing the evaluation team, and managing the review and finalisation of the evaluation report.

This is usually a subgroup of the steering group; it may be replaced by a single evaluation manager. For a large management group, it is a good idea to establish a smaller steering group that can take decisions quickly when necessary without the need to call a full meeting of the management group.

REFERENCE GROUP

Reference group
A group made up of primary stakeholders familiar with the local environment who can advise on practical issues associated with the evaluation and on the feasibility of the resulting recommendations.

Establishing a reference group is a good way to involve the primary stakeholders when an evaluation has been commissioned by headquarters.

PEER REVIEW GROUP

Peer review group
A group that advises on quality issues; usually made up of evaluators and other specialists chosen for their knowledge of evaluation, the region, or the type of intervention being evaluated.

A peer review group may also advise on the contextual analysis and methodology and comment on the draft report. It does not need to be large. It can be especially useful in evaluations that are managed by a general manager rather than an evaluator and when the evaluation team is unfamiliar with the region.

Keep the peer review and reference groups separate so that each can focus on its assigned task and avoid distraction. While their responsibilities may appear to overlap, they function better separately.

Advisory groups sometimes go by different names in different agencies. Large joint evaluations will benefit from all of them, but simpler arrangements are more suitable for smaller evaluations. Even the smallest evaluation can benefit from having a reference group of stakeholders and a peer review group or even a single reviewer.

3.7.2 Managing conflict

Disparity between the scale of the evaluation task and available resources may lead to conflict between the evaluation manager and the evaluation team. Many of these disputes stem from differences in understanding. An inception report can help lessen the risk of such conflict in the early stages, but conflict may also emerge at the report stage, leading to multiple revisions and increasing frustration on both sides.

One way of minimising disputes in contentious evaluations is to use a steering group to advise on the acceptability of draft reports (although the final decision remains with the evaluation manager). Steering groups can also help ensure the quality of the whole evaluation process.

Personality clashes are another potential source of conflict; evaluation is a field that attracts strong personalities. It may be useful to have a formal dispute resolution policy specifying what steps would be taken in the event of a dispute and identifying a neutral person who could arbitrate between the parties.

Good practice example 4

Establishing a dispute resolution procedure

The Tsunami Evaluation Coalition adopted a policy on resolving disputes within the evaluation teams, between a team and the steering committee, within the synthesis team, and between the synthesis team and the core management group. Under that policy, the primary author was responsible for managing relations within the team. When there was a serious dispute about a substantive issue between team members, or between a team member and the team leader, that the primary author was not able to resolve, the core management group was to ask the head of the ALNAP Secretariat for a report giving both sides of the issue. The head of the ALNAP Secretariat could either prepare such a report directly or contract an experienced evaluator to do so. The management group could then decide to ask the primary author to do one of the following:

- Present only one of the interpretations.
- Include both interpretations in the report.
- Include only one interpretation, but note that this interpretation was not unanimously held.

If members of the synthesis team were unhappy with the resolution of the problem, they had the right to have their name removed from the report.

Formal dispute resolution procedures are appropriate for any evaluation that needs them, not only for large evaluations.

3.8 Choosing an evaluation team

The decision on using an external, internal, or mixed evaluation team depends on the purpose of the evaluation. If the evaluation is principally for accountability, the evaluators should be external. Depending on the organisational context, it may be desirable to include one of two staff members on the team – perhaps from the agency's evaluation department if there is one, or, if appropriate, from the operational departments. These members should never be in the majority or take the leadership role; inclusion of any internal staff on the team can put the perceived independence of the evaluation at risk.

If the evaluation is principally for learning, either the evaluation team should include a majority of internal staff who are expected to do the learning, or a team of external evaluators should be engaged whose primary role is to facilitate staff learning.

Pros and cons of using internal and external evaluators are presented in **Table 6**.

Table 6: Advantages and disadvantages of using internal and external evaluators

Internal evaluators		External evaluators	
+	Benefits evaluators derive from learning and reflection during the evaluation process remain within the organisation.	+	They are often more objective.
+	They know the organisation and its culture.	+	They are less likely to have organisational bias.
+	They are known to staff.	+	They bring fresh perspectives.
+	They may be less threatening and more trusted.	+	They may have broader experience to draw on.
+	Findings and recommendations may be more appropriate for the organisation.	+	They may be able to commit more time to the evaluation.
+	Recommendations often have a greater chance of being adopted.	+	They can serve as an outside experts or facilitators.
+	They are less expensive.	+	They are not part of the organisation's power structure.
+	This option builds internal evaluation capability and generally contributes to programme capacity.	+	They can bring in additional resources.

Internal evaluators		External evaluators	
−	Their objectivity may be questioned.	+	They are more likely to be trained and experienced in evaluation.
−	Organisational structure may constrain participation.	+	They are regarded as experts.
−	Work commitments may constrain participation.	−	They may not know the organisation.
−	Their motivation may be questioned.	−	They may not know the constraints that will affect take-up of recommendations.
−	They may too easily accept the assumptions of the organisation.	−	The benefits they derive from reflection and learning during the evaluation process do not remain within the organisation.
−	They may not be trained in evaluation methods.	−	They may be perceived as adversaries.
−	Using insiders may reduce the evaluation's credibility outside the organisation.	−	They are more expensive.
−	They may have difficulty avoiding bias.	−	Hiring them may require time-consuming contract negotiations.
−	They may lack specialist technical expertise.	−	Using outsiders may make follow-up on recommendations less likely.

Internal evaluators	External evaluators
	They may be unfamiliar with the environment.
	They may be influenced by their desire for future work and thus be less independent than they appear.

Bias can also be an issue in evaluations.

Bias
Either an evaluator's prejudice, or systematic errors that arise because of the evaluation design.

Examples of the first type of bias include an evaluator's dislike for a particular agency or programme approach, and the temptation to repress hard-hitting findings out of fear of losing future evaluation work with the client.

An example of the second type of bias is poor coverage of less accessible locations due to a choice to visit only the most accessible locations.

3.9 Identifying external consultants

Getting the right consultant(s) for your evaluation is important and takes time. The further ahead you recruit, the more choices you will have.

The type of consultant varies with the evaluation task. When dealing with sensitive issues of partnership, you need consultants with an understanding of institutional relationships in general and partnerships in particular. For a real-time evaluation, you need consultants with sufficient operational experience to be credible to the staff in the field. Considerations in choosing consultants include the following:

- How well the consultant knows the organisation. A consultant who knows the organisation well will need to spend less time learning about it but may be less inclined to challenge its prevalent beliefs. How well the consultant knows the context. This is especially important for complex crises and conflicts. Depending on the evaluation, it might be more important to understand the context or the type of activity to be evaluated.
- Whether the consultant has the requisite skills and experience for the proposed evaluation approach.
- The overall team composition. Larger teams may offer more skills but may also cost more, require greater logistics support in the field, entail a bigger risk of internal disputes, and take longer to complete a report.

Another issue is whether to recruit consultants individually or as a team via a consulting company. Pros and cons of each choice are summarised in **Table 7**.

Table 7: Advantages and disadvantages of using individuals or a contractor's team

	Individuals		Contractor's team
+	You select exactly the people and skills that you want.	+	There is only one contract to manage.
+	You can ensure that all team members have something to offer.	+	You have more leverage on quality issues.
+	May be less complex than a contract.	+	Resolving team conflicts is the contractor's responsibility.
+	You have a direct relationship with everyone on the team.	+	You have to deal with only a few payments.
+	This option is usually less expensive.	+	The contractor assumes some of the risks.
−	Financial limits may prevent you from hiring the most skilled consultants.	−	This option is usually more expensive.
−	Resolving team conflicts is your responsibility.	−	The team may have only one senior expert supported by junior staff.
−	You have to provide human resources support to the team.	−	Some team members may never have worked together before.
−	You have to organise logistics in the field.	−	You may still have to deal with field logistics and other tasks if the contractor is not competent.

Individuals		Contractor's team	
–	You have to deal with many different payments.	–	Financial limits may prevent you from hiring a team in a timely manner.
–	You bear all the risks.		

Sometimes you can select consultants on the basis of your prior experience with them, but you won't always have this luxury. You can, of course, ask for short notes on the approach they would take, but a better option is to select consultants based on their previous work. Who wrote the evaluation reports that you consider of good quality?

The exact contractual relationship will depend on agency policy. A common practice is to link payments to stages in the evaluation, such as signing of the contract, acceptance of the inception report, completion of the fieldwork, and acceptance of various drafts and the final report.

Don't rely completely on a potential consultant's authorship of a previous evaluation – contact the evaluation manager and ask about the consultant's work record. The final report may have been the 19th draft and may have been completely rewritten by another team member.

3.10 Budgeting for an evaluation

The biggest cost driver for any evaluation is the total number of external consultant days. This is normally a product of the size of the team and the number of days of fieldwork, multiplied by the consultants' pay rates and daily subsistence rates. A budget can be broken down into preparatory, fieldwork, and post-fieldwork phases.

- **The preparatory phase** can include an initial meeting to discuss the evaluation, background reading and the desk study, an initial headquarters interview, and writing an inception report. The initial meeting may require 1 or 2 days including travel, but not all team members need to attend. Background reading, initial interviews, and the inception report can take from 1 to 5 days for a small evaluation and a month or more for a large, complex evaluation. A small evaluation (of a single activity at a single site by a single agency) will typically allocate 3 to 7 days per consultant for preparatory work.
- **Fieldwork** costs are a function of the length of the fieldwork and the cost of the consultants. As a general rule, quantitative methods demand more extensive fieldwork and larger field teams; qualitative methods require more highly skilled fieldwork; and quantitative methods require higher design and analysis skills. A small evaluation would typically have a week for observation and interviews at field sites sandwiched between half a week of initial briefing interviews and another half week of follow-up interviews and debriefing. More time is needed if there are multiple field sites to be visited, or if the team is expected to produce a draft report before leaving the country. A small evaluation would typically have 12–14 days for fieldwork including travel; a large evaluation may require months of fieldwork. A good rule of thumb is 1 week for every country visited plus at least 1 week for each major site (for example, province or district) to be visited. If the team is to engage in detailed beneficiary consultation, allow 2 or 3 weeks per site. Quantitative survey methods are more costly and should be budgeted according to the sample size and the time it will take to process and analyse the data.

- **Post-fieldwork** tasks consist primarily of collating and analysing the data collected and writing the report. They may also include debriefings, review cycles, and dissemination. Collating data and writing always take longer than the allocated time. The time needed varies with the complexity of the report and the amount of analysis needed. Debriefings can take 1–2 days, writing the report 1–20 days, and each review cycle 2–10 days. When there are many team members, a good rule of thumb is to add 5 days to the report preparation time to allow the team leader to incorporate their material. A small evaluation would typically allot about 7–12 days for post-fieldwork activities.

Common budget elements are outlined in **Table 8**.

Table 8: Cost elements for an evaluation

Budget item	Possible elements
Personnel	Staff pay and allowances, allowances for partners' staff and other staff
Consultants	Team leader, international consultants, national consultants, other consultants
Support staff	Pay and allowances for administration, background researchers, interpreters, drivers, security staff, and others
Travel	Visas, flights for evaluation team and accompanying staff, flights for briefings, internal travel by team and accompanying staff
Subsistence allowances	Accommodation and per diem costs for consultants
Data entry	Data input and cleaning of data to remove nonsense responses, such as someone recorded as both male and pregnant
Meetings and workshops	Venue hire, meals, and allowances
Report production	Copy editing, translation, artwork, graphic design, layout, printing, development of electronic media, distribution
Miscellaneous	Communications, mail and couriers, teleconferencing, licences and legal fees, security
Overhead costs	

3.11 Negotiating terms of reference

The evaluation design is expressed in the ToR, a key evaluation tool that provides stakeholders with a common understanding of the evaluation, its objectives and purpose, key questions, and other practical and strategic issues.

The ToR should be negotiated by the primary stakeholders of the evaluation. It enables the evaluation manager to define the evaluation task and tells evaluation team members what is expected of them. This common understanding of the task is essential to a successful evaluation (**Table 9**).

Developing the ToR through a scoping visit to the field can help ensure understanding between field staff and the evaluation office.

Table 9: Elements of the terms of reference

Element	Comments
Background	This requires only a paragraph or two; most of the evaluation team's contextual knowledge will come from background reading. Good sources for this include the introductions to project proposals or appeal documents.
Objectives and purpose of the evaluation	Is the evaluation mainly for learning or accountability (Section 3.1)? If both, which has precedence? For example, an evaluation may be intended to inform donors on how effectively their money was used (accountability) but might also examine which factors in project design and implementation led to the most effective projects (learning). All evaluations, whether primarily for accountability or learning, provide learning opportunities. It is useful to ask at this point if an evaluation is the best way of achieving the stated objectives. Would an audit better meet the accountability requirement? Would other processes (after action-reviews, workshops) better meet the learning requirement?
Scope of the evaluation	What sectors, geography, phase of the response, and time frame will the evaluation cover? Is it focused on the policy level? the programme level? on specific operations or processes? Is it a real-time evaluation, mid-term evaluation, or ex-post evaluation? Is it a single-agency evaluation, self-evaluation, or joint evaluation? Sometimes management or agency policy will define the scope. In other cases you may have to develop the scope through discussions with colleagues. Good ToR specify not only what is to be evaluated but also what is not to be evaluated, and why. For example, if an agency had recently evaluated its work on emergency water and sanitation, it would make sense to exclude this sector from the current evaluation.

Contents
1
2
3
4
5
6
7

Element	Comments
Expected use of the evaluation	How will the evaluation be used, and by whom? Being clear about this will help make the evaluation cost-effective. See Section **2** for tips on how to ensure that your evaluation has a utilisation focus.
Why the evaluation is being conducted, and why now	Are there any internal or external deadlines? Is the evaluation tied to decisions in a funding cycle or decisions about the agency's future strategy? Such deadlines and linkages should be clear if you have involved primary stakeholders from the outset, and should be made explicit in the ToR.
Relevant conceptual models and international standards	Is there a conceptual model that you want evaluation team members to use when designing their methodology and carrying out their analysis – for example, the livelihoods framework for complex humanitarian emergencies? What international standards are relevant to this evaluation – for example, Sphere standards or the People in Aid Code?
Main evaluation questions	What are the main questions you want the evaluation to answer? Which OECD/DAC criteria do you want to use, and how do they relate? *Tip:* ***Keep the number of evaluation questions short in order to keep the evaluation focussed.***
Methodology	This should not be a detailed statement, but rather an indication of any methodological preferences – for example, if you want the team to consult specific people, such as government officials or beneficiaries, or use particular methods, such as a formal survey. Normally, the team should develop the detailed methodology. If the evaluation manager specifies both product and process, this is likely to reduce the evaluation team's sense of ownership and of responsibility for the end result.

Element	Comments
Timetable	Specify dates for key deliverables and any deadlines. Ask stakeholders if there are any times that are not suitable – for example, the wet season, harvest season, elections, or administratively busy periods. *Tip: **Allow enough time for circulation of a draft report for comments (typically 2–3 weeks) and revision of the report (another 2–3 weeks).***
Roles and responsibilities	Specify who is responsible for providing transport in the field, making appointments, and other tasks. Check with your colleagues in the field to make sure that they can provide the necessary support to the evaluation team.
Management arrangements	Specify whether there will be advisory committees and what their roles will be.
Outputs	Outputs normally include an inception report, field debriefings (as a note or a presentation), a main report, an evaluation summary, and debriefings at general meetings (see Section **7.1**). Specify the length and the format if you have a preference. It is also useful to establish phased payments against specific outputs to encourage the evaluation team to maintain a goal-oriented approach. If you have a house style to which the evaluation report should adhere, or strong preferences about format and style, specify these to avoid having to reformat the report later. Specify the length of the report (number of words is a more precise measure and less susceptible to misunderstanding than number of pages) and what annexes you want. You may also want to specify the broad structure of the report, although it is good to give the evaluation team some flexibility in this regard.

Element	Comments
Risk management	Describe the risks and challenges that are expected to arise in the evaluation and how the evaluation manager proposes that they be dealt with.
Budget	Give an indicative budget for the evaluation.
Bid assessment	If there is no formal tender request document, you may want to include the basis on which bids will be assessed (what percentage of marks will go for price, team composition, methodology, and other criteria) and the deadline for the receipt of tenders.
Key references	You may want to attach a list of key references for the evaluation. Many of these will be internal documents, but check with primary stakeholders as to what other documents should form part of the initial reading set.

Contents
1
2
3
4
5
6
7

Key questions to ask during evaluation planning include the following:

- Are the key stakeholders 'on board' for the evaluation, and have they been adequately involved in drawing up and approving the ToR?
- Are the objectives and purpose clearly described?
- Is the scope of what is to be evaluated (and what is not) clearly stated?
- Is the intended use, along with any related timing constraints, clearly stated?
- Is the number of evaluation questions reasonable and manageable?
- Are any methodological requirements clearly stated?
- Are the roles and responsibilities, management arrangements, and desired outputs clearly stated?
- Are risks and anticipated responses to them identified?
- Is the budget clear?

Can this section be improved?

NOTES

NOTES

Please share your feedback:

Help us make the Guide more practical and user-oriented by sharing your feedback on its content and navigation. Feedback can include questions, suggestions, useful resources or practical advice from your own experience. You can share your comments and ideas by clicking on the feedback button at the bottom of any page. This will generate a confidential email to the EHA team in the ALNAP Secretariat. For more information on how you, your team or organisation can engage more actively in piloting the Guide contact us directly at **eha@alnap.org**

DOING AN EVALUATION OF HUMANITARIAN ACTION

Contents of this section

4.1 The inception phase 93

4.2 Leadership and teamwork challenges 108

4.3 Involving the affected population 111

4.4 Ethical issues 113

4

How to use this section

This section is targeted at both evaluators and evaluation managers. Section 4.1 covers the inception phase, during which the evaluators make sense of the evaluation task and plan their approach and methodology. Section 4.2 provides tips on teamwork and management to overcome common pitfalls in managing EHA. Section 4.3 provides guidance on consulting the affected population, and Section 4.4 addresses some ethical issues in doing EHAs.

Tables, figures and boxes

Table 10:	Elements of the inception report	95
Table 11:	Sample element of an evaluation matrix	97
Table 12:	Sample logical framework	100
Figure 6:	Sample results chain	101
Figure 7:	Logic model for a humanitarian evaluation	103
Table 13:	Inception report checklist	105
Figure 8:	The four elements of a successful evaluation of humanitarian action	108
Box 5:	Considering age and gender in consultations with the affected population	112
Figure 9:	Interviewee rights advice card	114
Box 6:	Payments to interviewees	115

4.1 The inception phase

The inception phase occurs prior to fieldwork; its output is the inception report. This can be relatively short, as in the case of the Haiti Emergency Relief Response Fund evaluation (Moriniere, 2011a) – which can be compared with the final evaluation report (Moriniere, 2011b).

Most ToR in humanitarian evaluation are fixed rather than negotiated between the evaluation manager and the evaluation team. The inception report allows the evaluation team to define a specific plan and get agreement on it from the evaluation manager, as well as to raise concerns and address any ambiguities in the ToR. For example, often the work implicit in the ToR would take a lot longer than is budgeted for; the inception report is an opportunity for the evaluation team to clarify what they can cover and what they cannot.

It is almost always worthwhile for the evaluation team to produce an inception report even if one is not required, as it helps the team to plan its work and eliminate possible misunderstandings between the team and the commissioning agency.

4.1.1 Preparing the inception report

The inception report provides the following opportunities:

- For the evaluation manager to assess how the team understands and plans to approach the evaluation
- For the evaluation team to turn the ToR into an achievable plan that is agreed with the evaluation manager
- For the evaluation team to seek clarification of the ToR and highlight tensions that need to be resolved by the commissioning agency (for example, conflicting expectations within the agency regarding the evaluation)
- For the evaluation team to plan its work in a coherent way
- For other stakeholders to receive a clear statement of intent by the evaluation team so that they can quickly flag any issues that they have with the proposed approach

The inception report should present a clear and realistic plan of work and timetable that takes existing constraints into account. The work plan should include the allocation of roles and responsibilities within the team, any deadlines for intra-team reporting, and detailed travel plans. The contents of the inception report vary depending on the context (**Table 10**). If timeliness is an issue that the evaluation is to examine, an outline chronology should form part of the inception report. The team can add more data to the chronology in the field.

The inception report should present proposed methodologies, including an initial priority interview plan for further interviews. It should acknowledge, and where possible specify the roles of, any advisory groups. In an annexe, it should include an interview guide and focus group topic list, if these activities are planned.

Contents
1
2
3
4
5
6
7

Table 10: Elements of the inception report

Element	Comments
Background	This section should summarise the context in which the evaluation is taking place. *Tip: **A chronology can be quite useful for setting out the background, and can be expanded for the main report with data gathered during fieldwork (see Section 5.3.3).***
Action to be evaluated	This section should show that team members understand what the action to be evaluated consists of. It may consist of a chapter describing the intervention with basic data collected by the team during the desk study, and may include a logic model (see Section **4.1.3**). Data tables may be included in an annex.
Purpose of the evaluation	This section should summarise team members' understanding of the purpose and objectives of the evaluation and the use to which it will be put, and explain how this has influenced their choice of methodology.
Methodology	This section should set out the methods the team proposes to use to gather and analyse data to answer the evaluation questions. It should indicate what sampling strategy the team intends to follow. It should also include (possibly as annexes) any interview guides, survey forms, or other data collection instruments to be used. It should clearly state the limitations of the proposed data collection methods, including the sampling strategy, and any limitations due to availability of resources.
Evaluation questions	This section is only necessary if the evaluation questions are reworked by the evaluation team – for example, to synthesise some questions to reduce the total to a manageable number.

Element	Comments
Evaluation matrix	This shows how the evaluators plan to answer each of the evaluation questions (see Section 4.1.2). It should reflect the methods set out in the methodology section. It may be presented in an annexe.
Detailed work plan	This specifies where team members plan to visit and when, and the days proposed for headquarters visits. It should also indicate which team member will be responsible for which task. Tip: In more insecure environments, it may be inadvisable to spell out the visit plan in detail in the inception report, because last-minute flexibility may be required.
Main report layout	This usually takes the form of a table of contents.
Interview targets	This provides a preliminary list of the people that the team intends to interview, or at least the types of people to be interviewed.
Outstanding questions and issues	This is an opportunity for the evaluation team to highlight ambiguities, areas of concern, or contradictions that they would like the commissioning agency to address and clarify before the next stage of the evaluation.

4.1.2 The evaluation matrix

A key part of most inception reports is the evaluation matrix, in which the team sets out a plan for answering each of the evaluation questions. (A sample matrix for one evaluation question is given in **Table 11**.) The evaluation matrix provides three pieces of information:

1. Questions to be answered – these are taken from the ToR (explicitly or implicitly).
2. How judgement will be formed – the criteria or indicators on which the answers will be based should be specified.
3. Expected information sources and methods – you should normally have multiple sources and methods for each evaluation question (Section **5.2.5**).

Table 11: Sample element of an evaluation matrix

Element	Evaluation question	How judgement formed	Likely sources and methods
Inclusiveness of country-level prioritization process	To what extent is the preparation of proposals at field level inclusive and transparent?	Evidence of inclusive and transparent processes at field level	Key informant interviews with Humanitarian Coordinators, NGOs, and agencies at headquarters and in the field
	To what extent are host governments engaged in proposal preparation? Is this level of engagement appropriate and sufficient? What are their perceptions of the process?	Evidence of engagement Expressed views of government officials and other humanitarian actors	Keyword analysis of proposals Key informant interviews with government, HC, and agency officials Survey of perceptions of the process

Element	Evaluation question	How judgement formed	Likely sources and methods
Inclusiveness of country-level prioritization process	To what extent are the greatest needs being prioritised, and what data are prioritisation decisions based on?	Evidence of prioritisation at country level Extent to which projects prioritised at the field level are funded	Key informant interviews with HCs, UN Office for the Coordination of Humanitarian Affairs country teams, agencies, and HCTs Comparison of funded applications with withdrawn or rejected applications

Source: adapted from the CERF 5-year evaluation.

An evaluation matrix shows how the team plans to answer each question, and reviewing it will allow the evaluation manager to see if the team has overlooked any major sources.

Evaluation managers may find it useful to prepare their own evaluation matrix and use it to check bids for the thoroughness of their proposed evaluation approaches. However, this matrix should not be shared with the team that wins the bid until after it has prepared its own version. The planning process is always more important than the plan, and the team should have the opportunity to engage with the evaluation questions thoroughly in its own way.

Contents
1
2
3
4
5
6
7
Feedback

4.1.3 Using a logic model

Logic model
A diagram presenting the programme theory (model of expected intermediate results and final outcomes) for an intervention.

ToRs frequently include a logic model for the operation (e.g., the CERF 5-year evaluation) or call on the evaluators to develop one. The growing popularity of theory-based evaluation and the growing attention to theories of change as frameworks for interventions are likely to increase the demand for logic models. Programme theory and theory-based evaluation are discussed in more detail in Section 3.2.5.

Logical frameworks are one of the simplest forms of logic model, with activities linked directly to outputs. **Table 12** presents a sample logical framework adapted from an approach originally advocated by NORAD (1999).

Table 12: Sample logical framework

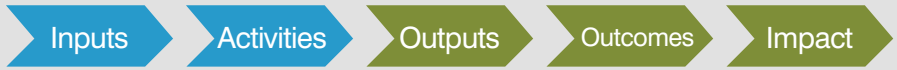
Results hierarchy	Indicators	Assumptions
Goal – the highest level objective towards which the project is expected to contribute	Indicators measuring progress towards the goal	Assumptions relating to the sustainability of the goal
Purpose – the effect that is expected to be achieved as a result of the project	Indicators measuring progress towards the purpose	Assumptions related to the achievement of the goal given that the purpose is achieved
Outputs – the results that the project management should be able to guarantee	Indicators measuring the extent to which outputs are produced	Assumptions related to the achievement of the purpose once the outputs are in place
Activities – actions undertaken to produce the outputs	Inputs: goods and services necessary to undertake the activites	Assumptions related to the production of outputs

Source: adapted from NORAD (1999).

The logical framework approach has been criticised as overly rigid, especially in fast-changing humanitarian contexts. Another issue is that the results hierarchy can be confusing, in that the goal of a project may be the same as the purpose of a programme containing that project.

NORAD eventually moved towards a more flexible approach using a results chain, as shown in **Figure 6**. (Names and precise definitions for the elements of a results chain can vary between organisations.)

Figure 6: Sample results chain



Evaluators are sometimes asked to develop a logical framework for a project that did not have one developed during the project design phase. In such circumstances, it is better to encourage the client to use a less rigid logical model such as a results chain.

The theory-of-change approach seems to be replacing the logical framework for some donors.

Theory of change
A description of the central mechanism by which change comes about for individuals, groups, and communities.

Sometimes the desired change might be expressed as the avoidance of death and suffering or the prevention or reduction of such contributory factors as acute malnutrition.

Theories of change may be supported by theories of action setting out what actions are taken to achieve the change. For example, if the theory of change is that improving household food security will reduce malnutrition in a crisis-affected population, the theory of action will set out how this will be done (for example, through cash transfers, food distribution, and market mechanisms). The recent OECD/DAC guidance on the evaluation of peace-building interventions includes an annexe on understanding and evaluating theories of change (OECD/DAC, 2012, p. 80).

Theory of action
A description of how interventions are constructed so as to activate the change described in the theory of change.

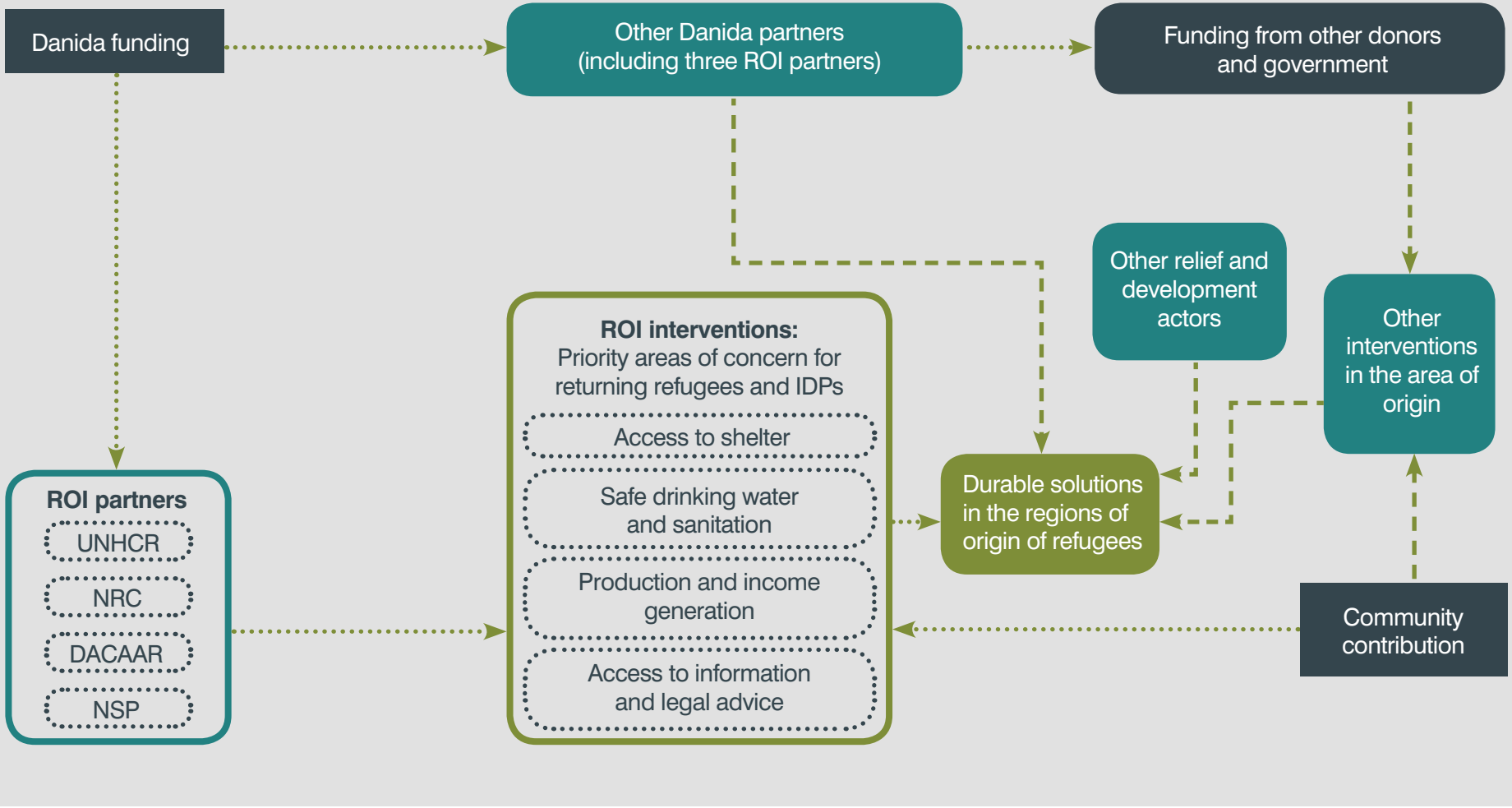
Programme theory may be explicit (where the theory is presented as a logic model), semi-explicit (where it can be constructed from a logical framework), or implicit (where there is no formal logical framework, but a logic model can be constructed from the way in which the project has operated and from interviews with key project staff).

Changes in the effectiveness of the intervention across the elements of a logic model can highlight the correctness of the assumptions made about the linkages between different elements.

If a broader conceptual framework for the sector exists, it can be useful to compare it with the logic model to see if there are major gaps in the logic model.

Logic models can also be developed from linkages implicit in programme documents. The logic model shown in **Figure 7** was developed for the evaluation of Denmark's ROI in Afghanistan. It was based on project logic that was implicitly suggested in various project documents.

Figure 7: Logic model for a humanitarian evaluation



Contents
1
2
3
4
5
6
7
Feedback

The Kellog Foundation has published a guide to developing logic models (WK Kellogg Foundation, 2004). However this is framed around quite simple results chains.

Models are convenient simplifications of complex reality. This simplification is their greatest strength, as it allows the manipulation of complex concepts and situations – but also their greatest weakness, as it may lead to you overlook a key factor. If you use a model in a validation, examine the extent to which the model fits the real situation you are examining and modify it as needed.

There are a number of software packages such as DoView that can facilitate the drawing of simple pipeline or outcome-chain logic models.

4.1.4 The desk review

Inception reports often include a preliminary desk review. Among other things, this may accomplish the following:

- Help you to select an appropriate conceptual framework if this is not already established.
- Assist in the development of a theory of change if this is not already clearly identified.
- Highlight key issues for further investigation during fieldwork.
- Allow mapping of the intervention by sector, geography, time, or other dimension.
- Identify key people to be interviewed.
- Identify potential targets for purposive sampling.

The resources consulted should ideally be presented in an annotated bibliography that demonstrates their relevance to the evaluation. Section 5.3 discusses desk review methods in more detail.

4.1.5 Assessing the inception report

The inception report allows the evaluation manager to see how the team understands the evaluation task and plans to approach it. It also allows stakeholders in the field to see how the evaluation team plans to do the work, so they can identify any issues or challenges they need to be aware of.

An inception report provides the evaluation manager with an opportunity to address problems with the team's understanding and approach before they become major issues. It should demonstrate the team's understanding of the context of the humanitarian crisis, the context of the response and

Contents

of the actions to be evaluated, the purpose and intent of the evaluation, and the concerns of stakeholders (**Table 13**).

Table 13: Inception report checklist

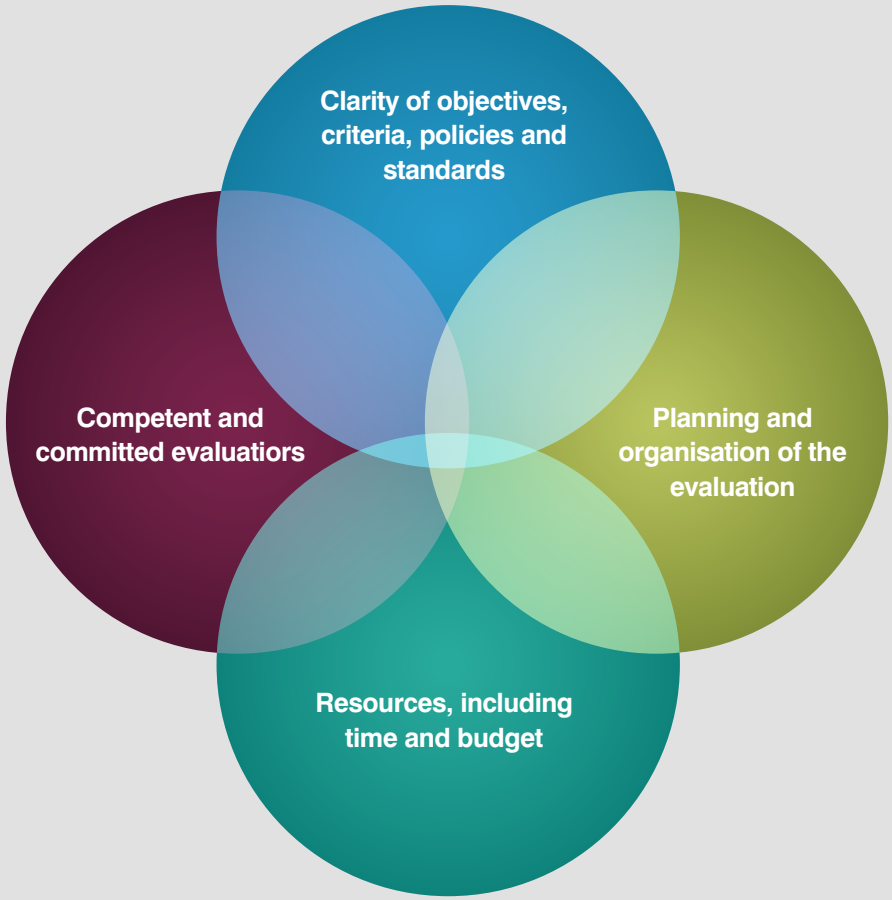
☐	Does the report demonstrate a clear understanding of the context?
☐	Does the report demonstrate a clear understanding of the intervention?
☐	Does the report demonstrate a clear understanding of the purposes and intent of the evaluation?
☐	Is there a clear plan of work?
☐	Is it realistic?
☐	Does the plan avoid conflict with any planned programme activities?

Feedback

☐	Does the plan identify the key dates for all activities and outputs?
☐	Are roles within the team clearly described and responsibilities assigned?
☐	Does the plan demonstrate a clear logical progression?
☐	Does the plan clearly identify and address likely constraints?
☐	Are all planned travel itineraries and dates given?
☐	Is the proposed methodology clear?
☐	Does the report show a clear understanding of the risks and limits inherent in the proposed methods?
☐	Does the evaluation matrix propose appropriate methods and judgement criteria (or indicators) to answer each evaluation question?
☐	Will the proposed methodology and amount and type of fieldwork allow the team to answer the evaluation questions as defined in the evaluation matrix?

☐	Has the team given adequate consideration to the role of any peer group, reference group, or other advisory group?
☐	Have the relevant advisory groups accepted the report?
☐	Does the report demonstrate the team's awareness of possible stakeholder concerns?
☐	Are the interview guide and other data collection tools presented in an annexe?
☐	Does the report indicate any areas of ambiguity or contradiction about which the team needs clarification before progressing?

Figure 8: The four elements of a successful evaluation of humanitarian action



4.2 Leadership and teamwork challenges

An EHA is like any other project – it needs to be well managed to be successful. There are potentially three levels of management: management by the evaluation department of the commissioning agency (the evaluation manager), management by the contract manager when a consultancy company is used, and management of the process and the team by the team leader. Each of these levels has different concerns. This section focuses on the management of the EHA process and of the evaluation team by the team leader. **Figure 8** summarises the four elements that are key to managing a successful evaluation.

Even when evaluators are competent, resources are adequate, and objectives are clear, good planning is still essential to the success of an evaluation. Almost all EHA is carried out against time constraints. Agreeing on ToR, recruiting a team, writing the draft report, circulating it, and incorporating reviewer comments almost always take longer than expected; fieldwork often suffers most when there are delays or cost overages. Careful planning can help avoid this.

Clarify from the start what support the client will provide to the evaluation team. Will the team have access to files, working space at headquarters and in the field, help with headquarters and field appointments, security services, transport in the field, and assistance with booking accommodation in the field?

The scale of the evaluation task determines the size of the team. However, as noted earlier, bigger teams bring problems and risks that may outweigh the benefits of their wider range of expertise. Bigger teams mean more work for the team leader, for example in managing assignments and collating inputs. They also pose a problem in insecure environments. Comfortable accommodation may be hard to find in some EHA settings, and larger teams may need to split up, making it difficult to have evening discussions.

Distribute templates for documentation like persons-met lists, bibliographies, and itineraries, formatted in the way you want to get the information. Getting inputs in a standard format minimises the work of converting and collating them.

EHA teams are usually assembled either by employing a consultancy company or by direct hire. In either case, the team leader may not have worked with all the other team members previously. Sometimes the team leader will be aware of an individual by reputation, but sometimes not.

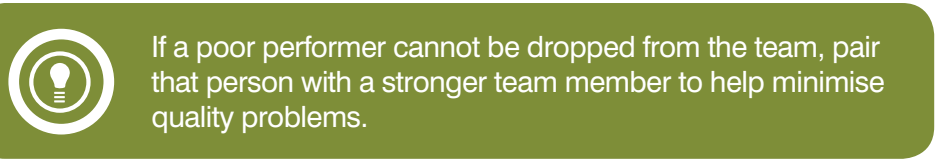
If you are working with some evaluators for the first time, and don't know them by reputation, organise the fieldwork so that you spend a day together at the start of the evaluation and can get a sense of their strengths and weaknesses, including any potential biases.

Even the best evaluators have off days, and previous good performance is not a guarantee of good performance on the current evaluation. Team members may be affected by concerns about their home life, the illness of a close relative, or other problems. The team leader should set deadlines for tasks and follow up immediately if the deadlines are not met. This can help to identify any performance issues quickly and allow corrective action.

Contents
1
2
3
4
5
6
7

Feedback

Sometimes problems occur within the team due to personality conflict, performance issues, or differing values. Some EHA environments can be stressful. Of these problems, performance issues are the easiest to deal with. However, by the time that it is clear that there is a performance issue, it may be too late to drop that person without abandoning the evaluation.



Large teams bring increased risk of conflict over findings, because of either differences in world view or differences between the areas that different team members are looking at. Tried and tested ways of managing large evaluation teams include the following:

- Use an evidence table (see Section 5.2.4) to keep track of evidence and emerging findings. Sharing this can keep team members informed and help build a common view.
- If you have a large team in a single country, build some time for reflection into the work programme so that the team can discuss emerging findings.
- Plan for team meetings, after fieldwork and before the completion of the first draft of the report, to discuss the evidence, findings, conclusions, and recommendations.

Sometimes a team member may be very concerned about a particular issue that, though relevant, is somewhat peripheral to the main thrust of the evaluation or is too narrow or too complex for inclusion in detail in the main report. In these circumstances, you can ask the team member to write an annex on the issue.

4.3 Involving the affected population

Consulting with the affected population has, for many years, been a weak part of EHA despite various efforts to increase accountability to this critical group of stakeholders, the people in whose name humanitarian action is undertaken[7]. If we are serious about listening to the affected population, then consulting them during an evaluation is essential.

For this to succeed, the affected population must be involved from the start, in determining the need for and purpose of the evaluation and helping to formulate the evaluation questions. This is rarely the case, and there may be other, more appropriate ways of ensuring that humanitarian action is accountable to the affected populations throughout the life of a programme or project (see for example HAP, 2010). However, the affected population should always have a voice in the evaluation, to communicate their views on the worth and effectiveness of the humanitarian action, unless there is a good reason not to – for example, in an evaluation that is focussed entirely on institutional processes rather than the effectiveness of a programme in alleviating suffering.

Paying attention to gender and age in consulting the affected population is critical, as these are key determinants of vulnerability, differential impact of the crisis, and different needs (see **Box 5**). While gender and age are always important, other factors such as disability, caste, or ethnicity may be important in particular contexts. Current UN evaluation advice recommends incorporating both a human rights and a gender equality perspective into all evaluations (UNEG, 2011). UNICEF has also published advice on designing and conduction equity-focused evaluations (Bamberger and Segone, 2011). See Section 5.4.2 for methods of consulting the affected population.

[7] Almost three-quarters of the EHAs that were assessed by ALNAP between 2001 and 2004 were judged unsatisfactory or poor in consulting with and encouraging participation by primary stakeholders, especially members of the affected population (Beck and Buchanan-Smith, 2008).

Box 5: Considering age and gender in consultations with the affected population

Natural disasters and armed conflict do not af¬fect everyone equally. Gender and age strongly influence how people are affected during and after a crisis. In the 2006 tsunami in south-east Asia, for example, mortality rates were much higher amongst women, children, and older people than amongst men. As a general rule, conflict-related emergencies kill more men, and natural disasters kill more women. In natural disasters, a larger gap between the socio-economic status of men and women is associated with a bigger difference between male and female mortality rates (Neumayer and Plümper (2007).

Gender and age can be key determinants of power dynamics and of vulnerability to gender-based violence. The views of women and young people may be drowned out by men in a community before, during, and after a crisis. Humanitarian relief projects must take into account ways that gender and age affect need and vulnerability. For example, in camps for refugees and displaced people, latrines must be placed in locations where women can use them safely without risk of attack; men who have lost their wives may have particular child-care needs.

The legal rights of men and women, for example to own land, may be different in different societies. This has sometimes been a major issue for households that become female-headed as a consequence of the crisis, for example after the Rwanda genocide.

For all of these reasons, sex and age should be considered carefully when collecting information from and about the affected population during an evaluation. The definition of older people and children should be adapted to the context in consultation with the local population.

Source: Mazurana et al. (2011).

4.4 Ethical issues

Each of us has many different roles with different responsibilities, and doing right in one role may clash with doing right in another. This is an ethical dilemma. For example, as employees, it is our responsibility to provide information to those who manage us when required, but this may conflict with our obligation to provide confidentiality to interviewees.

Various evaluation bodies have produced guidance on ethics. The American Evaluation Association's Guiding Principle for Evaluators (2004) emphasises systematic inquiry, competence, integrity and honesty, respect for people, and responsibility for general and public welfare. The Joint Committee's Program Evaluation Standards (1994) focus on utility, feasibility, propriety, and accuracy. The United Nations Evaluation Group's 2008 Ethical Guidelines concentrate on the intentionality of evaluation, the obligation of evaluators, the obligations to participants, and the evaluation process and product.

Familiarise yourself with any ethical policies established by your organisation and professional organisations of which you are a member, or which you have otherwise made a commitment to honour.

The main ethical issue in EHA lies with the potential for evaluations to do harm to informants and other stakeholders. Danger to informants is a critical issue, especially in complex emergencies where even talking to the evaluation team may put an interviewee at risk of detention or worse. Where identifying interviewees might place them at risk, terms like 'refugee one, refugee two' can be used. See Section 5.4.2 for further discussion of ethical issues in interviews with the affected population. Another way in which evaluations can do harm to stakeholders is the misuse of evaluations to improperly support decisions that have been taken in advance but are not supported by evidence.

Children are often overlooked in consultations. They have their own perspectives, and it is important to look at them when evaluating programmes that try to meet their needs. However, research with children raises particular ethical concerns. UNICEF has published a technical note on how to address these concerns (**UNICEF Evaluation Office, 2002**). Evaluations are a form of social research, which is normally conducted on the basis of informed consent. Interviewees should be informed of the interview ground rules. This is important given the huge power differences that can exist between interviewer and interviewee. It can be done with a simple card (**Figure 9**). Most humanitarian evaluations are conducted on the basis of a variant of the Chatham House Rule (**Chatham House, 2007**), under which comments by interviewees are not attributed to them either directly or indirectly.

Figure 9: Interviewee rights advice card

Your rights as an interviewee:
You have the right to terminate the interview at any time.
You have the right not to answer any question.
Nothing you say will be attributed to you directly or indirectly without your express permission.
The notes on this interview will not be shared outside the evaluation team.
If the client agrees, and we have your email address, **we will send you the draft for comments.**

Evaluations may also have negative outcomes for other stakeholders. If individuals are criticised in an evaluation, this may damage their career prospects. Similarly, evaluation findings that particular interventions are ineffective or inappropriate may lead to loss of employment or loss of benefits for those employed by or receiving aid from those interventions.

Sometimes such negative outcomes for individuals can be entirely appropriate. While evaluation challenges to poor-quality humanitarian actions may be unpleasant for those affected, those actions consume resources that could otherwise better prevent death and suffering.

Box 6: Payments to interviewees

Interviewees who participate in humanitarian evaluations are not normally paid. This raises the question of whether the relationship is exploitative, as the evaluator receives a direct benefit while the immediate benefit to the interviewee is unclear. This question most often arises in relation to focus group discussions, which require two hours or more of participants' time. In industrialised countries, focus group members are typically paid.

Most discussion of research ethics in developing countries concentrates on medical research (**Molyneux et al., 2012**), in particular on clinical research (**Benatar, 2002**). For EHA, the risk is that payments by one evaluation may cause harm by limiting participation in future evaluations. The following general guidance can be offered:

- Evaluators should not pay interviewees in cash or kind. However, it is appropriate to provide light refreshments and snacks. Participants in focus group discussions might also be given an agency tee shirt or other promotional gift as a thank-you for their time. However, no gifts should be given in insecure situations where they might either encourage participation or encourage targeting of the recipients.
- Evaluators should treat interviewees with respect, arrive on time for scheduled interviews, fairly represent the views of interviewees in their reports, and acknowledge their assistance.

Evaluators should not condemn named individuals or agencies for particular instances of bad practice, unless the evaluation has been able to review all of their work and can put that example of bad practice into context.

Neither should people or agencies be condemned unless they have had an opportunity to defend their actions. If evaluators come across actions that are clearly or potentially criminal, then reporting them directly to the evaluation manager and the line manager is more appropriate than including them in the evaluation report. Evaluators are not detectives and should leave the detailed investigation of potential criminal acts to those best equipped to do so.

Normally, the duty of an evaluator is to report wrongdoing to the evaluation manager, but there may be situations where this is not enough and where the evaluator is under a legal or ethical obligation to inform local authorities. In such circumstances, the evaluator should seek legal advice before reporting wrongdoing to anyone outside the organisation.

NOTES

NOTES

Please share your feedback:

Help us make the Guide more practical and user-oriented by sharing your feedback on its content and navigation. Feedback can include questions, suggestions, useful resources or practical advice from your own experience. You can share your comments and ideas by clicking on the feedback button at the bottom of any page. This will generate a confidential email to the EHA team in the ALNAP Secretariat. For more information on how you, your team or organisation can engage more actively in piloting the Guide contact us directly at **eha@alnap.org**

EVALUATION DESIGN AND METHODS

Contents of this section

5.1 Evaluation design 120
5.2 Research approaches 127
5.3 Desk review methods 142
5.4 Field methods 158
5.5 Learning-oriented approaches 176

5

How to use this section

This section can help evaluators and evaluation managers assess the pros and cons of different evaluation approaches and methods. Section 5.1 gives an overview of evaluation design. Sections 5.2 to 5.4 describe different data collection methods suitable to EHA, both desk- and field-based, with examples of how they have been used in different evaluation contexts and tips on the analysis process. Section 5.5 introduces some participatory processes that are particularly well-suited to learning-oriented evaluations.

Tables, figures and boxes

Figure 10: An implicit programme theory for humanitarian action 124
Figure 11: Factors affecting credibility and utilisation 127
Figure 12: Reliability and validity in quantitative research 134
Table 14: Questions to ask about qualitative research 136
Figure 13: Idealised chain of evidence 138
Table 15: Segment of a summary evidence table 141
Figure 14: Number of documents posted to ReliefWeb during a recent humanitarian response 145
Figure 15: Key events in a 2009 dengue fever outbreak in Cape Verde 148
Table 16: The Inter-Agency Standing Committee's gender analysis tool 156
Box 7: Using recall 159
Table 17: Challenges and ethics of consulting the affected population in a humanitarian crisis 162
Table 18: Ways of consulting the affected population 165
Table 19: Participatory rapid appraisal techniques 169
Box 8: Taking sex and age into account during field data collection 174

5.1 Evaluation design

Before selecting methods, the evaluator needs to consider the design of the evaluation. Many methods can be used with any evaluation design, but some methods are more suited to some designs than others.

Evaluation design
The research approach to be used in an evaluation.

Evaluation design is sometimes divided into three broad categories: experimental, quasi-experimental, and non-experimental (Bamberger et al., 2011a). However, this categorisation is somewhat limiting for humanitarian evaluation, and a recent DFID report (Stern et al., 2012) identified five types of impact evaluation design:

1. Experimental
2. Statistical
3. Theory-based
4. Case-based
5. Participatory

Process evaluations may use any of these designs but tend to focus on case-based and participatory approaches.

The different designs use different bases for inferring the causes of outcomes. Experimental designs involve comparison with a counterfactual.

Counterfactual
The condition that would have prevailed if a given humanitarian action had not taken place.

Statistical designs involve correlations and =different procedures to control confounding variables. Theory-based approaches involve identification or confirmation of causal chains or mechanisms. Case-based approaches establish similarities between cases or groups of cases. Participatory designs involve validation by participants of the causes of outcomes.

Bamberger et al. (2011a, p. 31) take a different approach, categorising impact evaluations based on when the fieldwork takes place (before, in the middle of, or after an intervention) and on whether there is

a comparison group. The majority of humanitarian evaluations take place during or after an intervention, and the use of comparison or control groups is rare. Process evaluations take place during implementation.

5.1.1 Experimental designs

This category includes randomised control trials (RCTs), quasi-experimental designs, and natural experiments.

RANDOMISED CONTROL TRIALS

Randomised control trial
An experiment in which subjects are randomly assigned to treatment and control groups.

All true experimental designs take the form of RCTs. With random selection, any individual has an equal chance of being selected for the treatment or control group. For humanitarian evaluations, families or communities may be selected rather than individuals. The RCT is the strongest and most robust evaluation design in terms of eliminating selection bias.

Selection bias
The risk that those who participate in an evaluation or benefit from a humanitarian intervention differ in some unobservable way from those who do not.

The strongest form of RCT is the double-blind trial, where neither the individual nor the researcher is aware of whether the individual is in the treatment group or not. One example of the use of a double-blind trial in the humanitarian sector was the efficacy testing of probiotic and prebiotic supplements with severely malnourished children in Malawi (Kerac et al., 2009).

RCTs face a number of potential problems in humanitarian contexts. First is the ethical issue of assigning assistance to the affected population on a random basis rather than solely on the basis of need. This is normally addressed by providing different but similar assistance to treatment and control groups. In the case of the Malawi trial (Kerac et al., 2009), all the malnourished children received similar amounts of therapeutic food, with the trial item included in the food for the treatment group

only. In evaluations of cash transfers versus food, the World Food Programme provided either one or the other to randomly selected treatment and control groups in Sri Lanka (Sandström and Tchatchua, 2010) and Malawi (Audsley et al., 2010).

Good practice example 5

Building the experiment into the intervention from the start

The International Food Policy Research Institute conducted an evaluation that compared food and cash distribution by the World Food Programme in Sri Lanka after the 2004 Indian Ocean earthquake and tsunamis. The team randomly distributed either food aid or cash to each recipient. Because both groups received desirable types of assistance, the project avoided the ethical and practical problems that arise when one group is assisted in a humanitarian crisis and another is not (Sandström and Tchatchua, 2010).

Technical problems with RCTs include the need to deliver the treatment in a standardised way, the difficulties of maintaining stable membership in the treatment and control groups, and the expense and effort needed. These constraints led a review on evaluations of health promotion by the World Health Organization in Europe to find that the use of RCTs for this was 'in most cases, inappropriate, misleading and unnecessarily expensive' (WHO Regional Office for Europe, 1998, p. 5). RCTs also require that the evaluation be designed from the start of the project with random selection of control and treatment groups. For all of these reasons, RCTs are rare in humanitarian evaluation.

Good practice example 6

Blinding to reduce the risk of bias

In the IRC evaluation of community-driven reconstruction in Lofa County, communities were assigned to treatment and control groups at the start of the project through a lottery (Fearon et al., 2008). One part of the research was conducted by an NGO that had not been involved in the programme. The NGO conducted the tests of community function without being told which group each community was in.

QUASI-EXPERIMENTAL DESIGNS

Quasi-experimental design

An evaluation design in which a comparison group is selected separately from the treatment group.

Quasi-experimental designs depend on a number of techniques to match the comparison group with the treatment group. They are susceptible to selection bias, as the comparison group may be significantly different from the control group. An otherwise excellent evaluation of UNHCR's cash grants in Burundi attempted to compare the situation of recent returnees who had received grants with earlier returnees who had not, but could not do so effectively given the other differences between the groups (Haver et al., 2009).

NATURAL EXPERIMENTS

Natural experiment
A situation in which membership in a treatment or comparison group is determined by factors outside the control of the project.

Natural experiments can occur where one group is affected or assisted and another similar group is not, or the two groups are affected or assisted differently. For example, a review of the impact of Hurricane Mitch on child survival compared children in areas hit by the hurricane with those in areas which were not hit by it (Baez and Santos, 2007). Natural experiments can also occur when conditions vary over time, as in the case of refugee influxes into western Tanzania (Maystadt, 2011).

Natural experiments can be very simple, like one in Niger comparing daily wages with piecework rates (Kevlihan, 2010). In another example, some local government jurisdictions were more prepared for the 2006 Yogyakarta earthquake as they were already on alert for an eruption of Mount Merapi. The Fritz Institute found high levels of satisfaction with the emergency response in those areas (Bliss and Campbell, 2007).

Brancati (2007) used a natural experiment design in her paper on the impact of earthquakes on interstate conflicts. In joint evaluations, it may be possible to compare groups who were assisted by different agencies using different approaches.

Evaluators should always be on the lookout for groups that did not receive assistance or that received different types of assistance during a humanitarian response. Sometimes such groups (provided that they are similar in all other respects) can form a natural experiment. Surveys of affected populations may reveal such natural experiments.

5.1.2 *Statistical designs*

Statistical designs rely on statistical tools to compare different groups. Statistical designs, especially econometrics, are widely used for research on conflicts (Do and Lyer, 2006) but have been little used in EHAs. They are quite demanding in terms of the statistical skills needed to conduct robust studies.

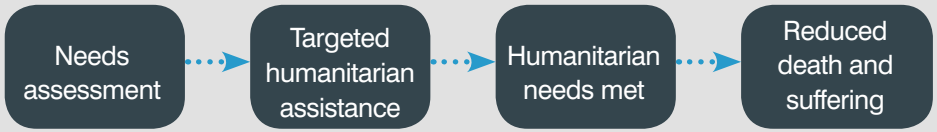
The econometric studies conducted by the World Bank are an example. Additional examples are the use of linear regression analysis to determine the importance of different factors for resilience after the Haiti earthquake (Tulane University's Disaster Resilience Leadership Academy and University of Haiti, 2012) and the use of an econometric model in the US State Department's evaluation of returnees to Burundi to compare returnees with non-returnees (Telyukov et al., 2009).

Interrupted time series are often used for research related to disasters, such as studies of the impact of the 1999 Taiwan earthquake on suicide (Yang et al., 2005) and the impact of the 1976 Guatemala earthquake on the division of labour. However, interrupted time series have been little used for humanitarian evaluations, probably because of the difficulty of separating the impact of the disaster from the impact of the humanitarian intervention.

5.1.3 *Theory-based designs*

As noted in Section 3.2.5, theory-based evaluations are based on a programme theory. Currently, few humanitarian evaluations are explicitly theory based, although a great many are based on an implicit theory of humanitarian action such as that shown in **Figure 10**, or on the logical framework for the intervention.

Figure 10: An implicit programme theory for humanitarian action



Contents
1
2
3
4
5
6
7

Recent evaluations that identify themselves as theory-based include a multi-donor evaluation of support to South Sudan (Bennett et al., 2010), an evaluation of resilience in Haiti (Tulane University's Disaster Resilience Leadership Academy and University of Haiti, 2012), and UNICEF's evaluation of the Early Childhood Development programme (Boller et al., 2011). Although humanitarian evaluations may be partially theory-based, this is not often emphasised or even clearly identified in the evaluation reports.

The growing popularity of the theory-of-change approach and the increasing emphasis on programme theory mean that explicitly theory-based designs are likely to become more popular.

5.1.4 Case-based designs

A DFID study (Stern et al., 2012) uses the term case-based to refer to a much wider range than is traditionally understood by *case study*.

The term *case-based* includes all evaluations that focus on specific cases rather comparing and contrasting a number of different cases as case-studies traditionally do.

Many current humanitarian evaluations fall into this category. Case studies comparing interventions across several countries are common for large humanitarian evaluations. For example, the CERF evaluation (Channel Research, 2011) examined the use of the fund across 16 counties (6 through country visits and 10 through desk studies).

5.1.5 Participatory designs

One aspect of participation is participation by evaluation stakeholders. Some participatory evaluations include the affected population not just as sources of information but as active elements of the evaluation. Such participants are also seen not as passive recipients of assistance but as active players in the response. It is rare to see this, however, in humanitarian evaluations, which sometimes do not even use the affected population as sources of information. Consultation with affected populations continues to be a problem in humanitarian action (Cosgrave et al., 2010, p. 4), including in EHAs. The 2002 ALNAP Annual Review (pp. 178–179) found that only 27 per cent of the evaluations reviewed had consulted with the affected population and had described this in a satisfactory manner.

Another aspect is participation of those being evaluated. The real-time evaluation of the Pakistan floods held stakeholder workshops to validate the team's findings and conclusions and to prioritise the team's recommendations (Polastro et al., 2011b). The 2007 NGO joint evaluation of the Yogyakarta earthquake response (Wilson et al., 2007) went even further in holding a multi-stakeholder event to review and amend the conclusion and make recommendations.

5.1.6 One-off versus longitudinal designs

Most humanitarian evaluations involve only a single period of fieldwork. Evaluations that revisit the same subject over time, called longitudinal studies, are common in development contexts but rare in humanitarian evaluation. They are particularly good at identifying changes over time. One example of a longitudinal study in a humanitarian context is the second evaluation of the linkages between relief, rehabilitation, and development following the tsunami (Brusset et al., 2009). Others include the World Vision evaluation of tsunami-affected communities in India, which revisited the same communities after 2 years, and a review of urban livelihoods in Afghanistan during which the team visited the subject households every month for a year (Beall and Schutte, 2006).

Although longitudinal evaluations are rare, good monitoring is longitudinal, and the differences between earlier and later monitoring reports may provide evidence of longitudinal change.

5.1.7 Choosing a design

The choice of evaluation design will be determined by the context and your objectives. Each design has different advantages and disadvantages. Options may be limited by your time window, budget, and the skills available.

5.2 Research approaches

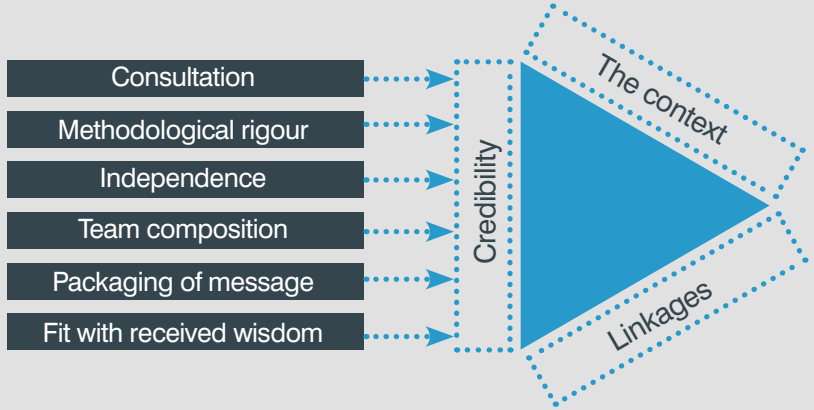
5.2.1 Methods

EHA employs both quantitative methods, focused on numerical data, and qualitative methods, focused on non-numerical data. An evaluation may include elements of both. For example, structured surveys are a quantitative method in which the questions are fixed and standardised and asked of a randomly selected sample of the population. Other types of surveys include open-ended questions that are not coded by the enumerator; these are essentially qualitative. Semi-structured key informant interviews are an example of a qualitative method. However, the data from qualitative research may be recorded in numerical format.

Development evaluation has been experiencing a replay of the 'paradigm wars' of the 1970s and 1980s (Denzin, 2010), with some arguing that quantitative methods are the only research approach producing objective evidence and others strongly challenging this view (Picciotto, 2012). Many references to quantitative methods of evaluation are not to the category in general but specifically to the use of surveys (typically face-to-face).

The important thing is to use methods that are feasible in the context, well suited to the research question, and seen as credible by stakeholders. Credibility, or the perception of quality, heavily influences utilisation (Court and Young, 2003); factors that affect credibility include rigour, independence, and consultation (**Figure 11**).

Figure 11: Factors affecting credibility and utilisation



Clearly, evaluators need to state what methods they have used, but it is also good practice for them to indicate what standard or guide their particular application of the method is based on. This not only provides readers with more information with which to assess the quality of the evaluation but also helps others who might want to use the same method in their own work.

Current guidance for evaluation in international relief and development settings favours the use of mixed methods in a range of designs (Bamberger et al., 2010), even for impact evaluation (Leeuw and Vaessen, 2009; Stern et al., 2012), in order to provide more coherent, reliable, and useful evaluations (Bamberger, 2012; Adato, 2011). Mixed methods are at their strongest when they support each other. This can happen, for example, when quantitative research (a survey) is based not only on the initial baseline but also on an initial round of qualitative research that identified key variables and hypotheses. Once the survey is conducted, further qualitative research can explore questions raised by the survey, and use the survey to assist purposive sampling, so that key categories are covered (Adato, 2011).

An alternative approach, with qualitative fieldwork sandwiched between two periods of quantitative work, is advocated by Bolton et al. (2007). They argue for qualitative interviews to draw up free lists of impacts between the pre- and post-intervention quantitative surveys, and give examples of where such qualitative fieldwork identified unexpected impacts.

Unfortunately, for most humanitarian evaluations, time pressures mean that quantitative and qualitative research take place in parallel, severely limiting the opportunities for synergy between the methods. An InterAction paper on mixed methods in impact evaluation (Bamberger, 2012) identifies a number of other possible sequences, but none offers all of the synergies that the qualitative-quantitative-qualitative sequence offers.

5.2.2 *Sampling*

It is normally not possible to interview everyone in the affected population or to assess every element of a programme. We compensate for this by working with a representative sample of the population.

Sampling
The selection of people or cases from a larger population for detailed research.

One approach to sampling is to draw a random sample from the population. Random sampling is not confined to quantitative methods, but is critical if you want to generalise the results from the sample to the whole population.

Random sample
A sample drawn from a population in which each member has an equal chance of being included in the sample.

Normally we don't have full details on the population and use some proxy instead as a sampling frame, and draw the sample from this. Beneficiary lists might be used as a sampling frame in humanitarian surveys, but such lists are often out of date, inflated, or incorrect. In a well organised refugee camp, it might be possible to use a master shelter list as a sampling frame.

True random sampling is often very difficult in the disruption typical of humanitarian contexts, so forms of pseudorandom sampling may be used. A common type of pseudorandom sampling may be selecting every fifth house, with the respondent selected using a pre-devised grid. Computers can be used to generate random numbers (technically they are pseudorandom as they are generated by a mathematical rule) to aid selection.

Contents
1
2
3
4
5
6
7

Good practice example 7

Using lists of affected sites as a sampling frame

The 2006 surveys for the Tsunami Evaluation Coalition used lists of affected villages and camps for internally displaced people as the initial sampling frame. Sites were then selected from within the sampling frame, and beneficiaries were selected pseudorandomly for interviews at these sites (**Brusset et al., 2006**).

If you need to do a random sample of a project list, a spreadsheet programme can generate random numbers that you can use to select projects. Often you may need to stratify the sample in some way to ensure that a cross-section of project types is represented.

Good practice example 8

Using random numbers to generate a stratified random sample

In a 2011 CERF evaluation, grants were sampled from 16 countries over a 5-year period and two different funding windows. The team wanted a random sample but also wanted to ensure that each year, organisation, and funding window was well represented in the sample.

Each project was assigned a pseudo-randomly generated number. Then, the projects were sorted by country, year, funding window, and recipient organisation. Each country had a nominal sampling ratio to select between 20 and 30 projects per country. The nominal sampling ratio was then applied to calculate how many projects should be selected for each organisation (for a particular country, year, and funding window), with a minimum of one. This number of projects was then selected for each organisation (for a particular country, year, and funding window) with the project with the highest random number being picked first.

CLUSTER SAMPLES

Some surveys seek to avoid the need for a sampling frame by using a clustered sampling approach.

Cluster sample
A sample that is selected by first dividing the population as a whole into clusters (usually geographical areas) and then selecting from within the clusters.

Clusters do not have to be geographical areas; for example, they could be schools or camp sections. Typically, the selection within the cluster is made by using some systematic pseudorandom rule from within the cluster (for example, the nearest house to the selected point followed by every nth house in the same direction). Recent research has also suggested that random selections can be made through the use of aerial or satellite photos (Shannon et al., 2012).

Cluster samples are commonly used for nutrition surveys with a sample of 30 drawn from each of 30 clusters. Of all of the random sampling techniques, cluster sampling has the highest sampling error (that is, the highest risk of not being representative of the whole population). Sample sizes have to be bigger than random samples drawn from the whole population in order to achieve the same precision.

SAMPLE SIZE

The sample size needed to be able to generalise the findings from the sample to the whole population in a statistically reliable way depends on the following:

- The closer this proportion of the population that has the factor you are interested in – the closer the proportion is to 50 per cent, the larger is the sample needed to estimate it with precision.
- The variability of this factor in the population (usually expressed as the standard deviation, estimated by experience or previous surveys) – the greater the variability, the larger the sample size needed.
- The margin of error or precision that is needed – tighter precision demands a larger sample size.
- The population size – this is important only if the sample size is more than 5 per cent of the population size.

Large-scale nutritional surveys typically use clustered samples with 30 clusters of 30 samples (30 x 30), and the samples for the Tsunami Evaluation Coalition Survey used about 1,200. Variants such as 30 x 7, 33 x 6, and 67 x 3 have been suggested as alternatives where high levels of prevalence are expected, such as for vaccination coverage and relief kit distribution (Bilukha, 2008).

Calculating sample size is technically challenging, and it is best to seek assistance from a professional survey statistician to ensure that your survey will yield the information you need.

For budgeting purposes, if you have no better information, it is probably safe to assume you will need a sample of about 1,000–1,200 people.

Samples for qualitative methods are usually relatively small, and are purposive rather than random. The advantage of a purposive sample is that you can concentrate on those cases that are of greatest interest.

Purposive sample
A sample that is selected in a purposive way.

Typically, purposive samples are selected to include people thought to be best able to provide the data needed (for example, key informants). The strategy behind purposive samples distinguishes them from convenience samples.

Convenience sample
A sample that is chosen on the basis of convenience rather than of any underlying strategy.

Convenience sampling is the weakest sampling strategy as those convenient for sampling may be atypical of the affected population. Those close to the road are much easier to sample, but they also receive more assistance than those who are harder to access, as UNICEF's East Timor Education Evaluation found (Tolani-Brown et al., 2010, p. 51).

One purposive sampling technique that is widely used in qualitative research is snowball sampling.

Snowball sample
A sample that is expanded by asking the initial interviewees to recommend others they believe would be useful sources of information.

Snowball sampling is often used to establish a sample for semi-structured interviews with key informants. It can also be used for populations that would be hard to sample in any other way.

Good practice example 9

Snowball sampling with telephone surveys
The 2009 evaluation of the use of cash grants by UNHCR in Afghanistan used snowball sampling in telephone interviews in Iran. Each household telephoned was asked to identify others who could provide useful information. It would otherwise have been very difficult to sample this group effectively (Davin et al., 2009, p. 13).

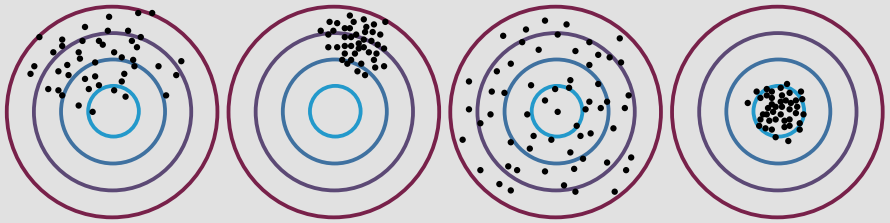
Whether you used quantitative or qualitative methods or a mixture of both, make a clear statement about what sampling strategy you used and what implications this has for your research. Being open about any constraints increases your credibility.

5.2.3 Controlling research quality

The first test of an evaluation's quality is the test of worth; and this lies at the heart of any approach that is focused on the likely utilisation of the evaluation (see Section 2.4). Evaluation methods must provide a reliable and valid account. Hammersley (1992, p. 69) defines an account as valid or true if it 'represents accurately those features of the phenomena that it is intended to describe, explain or theorise.' EHA uses mainly qualitative methods and to a lesser extent quantitative methods. One of the biggest differences between the two is the way in which they establish the extent to which they are a true reflection of the underlying reality.

Quantitative methods strive to be both valid (reflecting the true underlying condition) and reliable (returning the same result in repeated trials of the same situation) (**Figure 12)**. Standardised measurement approaches and question sets are used to ensure the reliability of quantitative approaches. Thus, enumerators are told to use a certain specific wording and not to offer any supplementary explanation.

Figure 12: Reliability and validity in quantitative research



Pattern reflects how well findings conform to the underlying reality.

The light-blue centre circles are realities and the dots are data points.

Quantitative methods depend on statistics to establish the validity of claims. If the sample is randomly selected and meets certain statistical tests, then the results from the sample can be generalised to the whole population. Quantitative data are normally perceived as authoritative and scientific. The limits of qualitative research are often much more apparent to stakeholders than the limits of quantitative research. This perception springs from a lack of understanding of the limits of quantitative data. For example, in a formal survey, the responses that enumerators elicit from interviewees depend upon who they are and how they are perceived by the interviewee. Such effects are very difficult to control for, and pre-tests are rarely done on a scale sufficient to highlight such effects.

The enumerator is one of the principle sources of error in structured face-to-face interviews (O'Muircheartaigh and Campanelli, 1998, p. 63). There is a rich literature on how responses can be influenced by characteristics of the enumerator, including gender (Flores-Macias and Lawson, 2008; Kane and Macaulay, 1993; Webster, 1996), status (Riessman, 1977), race (Hill, 2002), ethnicity (Webster, 1996), general social distance (Dohrenwend et al., 1968), and even whether the enumerator is thin or fat (Eisinga et al., 2011).

Always look at quantitative research results carefully. Pay attention to the confidence interval. For example, a newspaper headline may say that support for a political party has fallen by 4 per cent. However, such polls normally have a 95 per cent confidence limit of plus or minus 3 per cent, so it is possible for a party's 20 per cent support to show as 18 per cent in one survey and 22 per cent in another without having changed at all. In EHA reports, look at how the data were collected and how the survey dealt with common constraints in humanitarian crises such as the reluctance of interviewees to reveal personal information.

How representative was the sample?
In your own reporting, make sure that you indicate the confidence interval, and if appropriate, the statistical significance for your findings.

Qualitative methods need to be assessed in a different way; Mays and Pope (2000) suggested the questions summarised in **Table 14**.

Contents
1
2
3
4
5
6
7

Table 14: Questions to ask about qualitative research

Question	How to establish this in EHA
Worth or relevance – Was this work worth doing? Has it contributed useful knowledge?	Utilisation of the evaluation would establish its worth.
Clarity of the research question – Was the research question clear – if not at the outset, at least by the end of the research process? Was the researcher able to set aside his or her preconceptions?	The inception report serves as a key tool for establishing the clarity of the evaluation questions.
Appropriateness of the design – Would a different method have been more appropriate? For example, if a causal hypothesis was being tested, was a qualitative approach appropriate?	The methodology section should indicate why the particular mix of methods was chosen.
Context – Is the context or setting adequately described so that readers can relate the findings to other settings?	The evaluation report should include a chapter setting out the context.
Sampling – Did the sample include the full range of possible cases or settings so that conceptual rather than statistical generalisations could be made – that is, did it entail more than convenience sampling? If appropriate, were efforts made to obtain data that might contradict or modify the analysis by extending the sample (for example, to a different type of area)?	The methodology section should describe the sampling approach used and its expected effect on the data, as well as practical constraints faced by the evaluation team, such as lack of access to certain areas or population groups.

Question	How to establish this in EHA
Data collection and analysis – Were procedures systematic? Was a sufficient audit trail provided so that someone else could repeat each stage, including the analysis? How well did the analysis incorporate all observations? To what extent did the analysis develop concepts and categories capable of explaining key processes or respondents' accounts or observations? Were the interpretation of the data, and the theory on which it was based, easy to understand? Did the researcher search for disconfirming cases?	The methodology section should include examples of any data collection instruments used and explain the analysis process. The report should show a clear linkage between evidence, findings, conclusions, and recommendations. The time scale of humanitarian actions often requires evaluators to use shortcuts (such as working directly from handwritten notes rather than using elaborate analysis procedures with coding software), but the methodology should remain as rigorous as possible and should be clearly described in the report.
Reflexivity – Did the researcher carefully assess the likely impact of the methods used and his or her own influence on the data obtained? Did data included in the reports provide sufficient evidence for readers to assess whether analytical criteria had been met?	The evaluation should include brief evaluator biographies to enable readers to judge what weight to put on the analysis and potential bias. The methodology section should detail any limitations of the methodology.

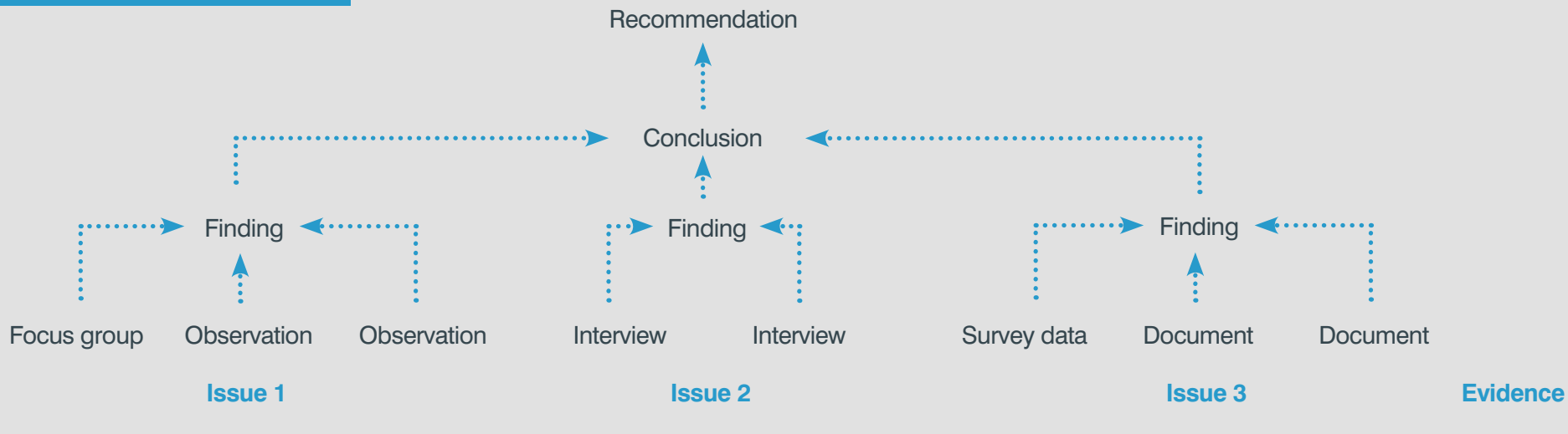
Sources: Mays and Pope (2000), (ALNAP, 2005).

In short, evaluations should provide the rationale for choosing the methods used and a full description of the sampling strategy, the constraints, limits imposed by non-response, and other potential sources of bias.

5.2.4 Building an evidence base

The most common criticism of evaluations is that the recommendations are not based on the conclusions, the conclusions are not based on the findings, and the findings are not based on the evidence. (Of course, this criticism may also be driven by a dislike of the conclusions and recommendations.) **Figure 13** shows the way in which a recommendation should be based on evidence – ideally, multiple pieces of evidence collected using different evaluation methods.

Figure 13: Idealised chain of evidence



A piece of evidence is a data point collected from an observation, interview, document, or other source. Not all pieces of evidence are of equal weight; bear this in mind if there are conflicts in the evidence. Findings are the inferences drawn from the evidence. Conclusions are higher-order inferences drawn from the findings after some thinking about what they mean. Recommendations are suggestions for action to address gaps emerging in the conclusions and to improve performance.

You can strengthen your evidence chain by using a simple table to record each piece of evidence for each evaluation issue. This could be a spreadsheet with columns listing the issues, the evidence, its sources (for example, interviews, documents, observations, and surveys), the date entered (to facilitate collation), and the initials of the team member entering it.

Using an evidence table allows you to see for which issues you lack enough evidence to reach a finding. Sometimes it may be possible to redirect the data collection effort to concentrate on these issues.

Structure your evidence table (and all your data collecting instruments) around the expected structure of the evaluation report. Immediately relating the data collected to their ultimate use avoids wasting time at a later stage in rearranging the material.

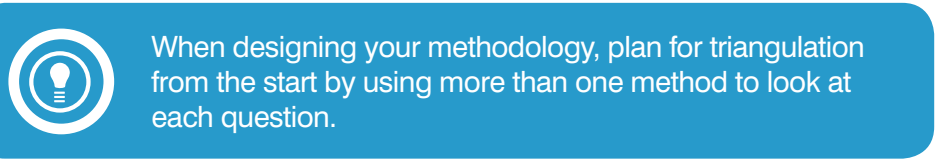
In a multi-country or multi-site study, schedule enough time after the first country or site visit to fine-tune your data collection instruments.

5.2.5 *Using triangulation*

Triangulation
Comparing data from different sources to see whether they support the same finding.

Triangulation is a key technique for ensuring accuracy and reliability in qualitative and mixed-methods research.

- **Method triangulation** compares data generated by different research methods – for example, observations and group interviews.
- **Source triangulation** compares information obtained from different sources – for example, key informants at headquarters and in the field.
- **Evaluator triangulation** compares information collected by different evaluators – for example, different team members.
- **Analytical triangulation** compares the results of different analytical techniques – for example, a study focused on the number of references in official documents to a particular issue (content analysis) and one focused on the level of funding for the same issue (numerical analysis).



5.2.6 From data collection to analysis

Using an evidence table also facilitates analysis as it organises all of the evidence for a particular issue (**Table 15**).

Table 15: Segment of a summary evidence table

Issue	Evidence	Finding	Conclusion	Recommendation
Volunteer skill level	1. Trainer in mock drill five did not push head back far enough to clear tongue for airway in demonstrating mouth-to-mouth resuscitation. 2. One-third of rope-joining knots shown in mock drills were inappropriate and dangerous. 3. The resuscitation method demonstrated most frequently is no longer regarded as very effective.	Volunteer and trainer skills need polishing.	Training needs a quality control mechanism.	The agency should set up a quality control mechanism within 6 months to ensure that training is of good quality and that skills are kept up to date.

When all of the evidence for a particular issue is listed together, it is easier for the evaluator to consider it thoroughly before writing on that issue. It is also easier for multiple team members to participate in the analysis.

When an evidence table contains information from interviews, it cannot be shared outside the team without breaching confidentiality.

5.2.7 *Counterfactuals*

Quantitative evaluations typically use comparison with a statistically valid counterfactual to establish what effect an intervention had. This is particularly the case with impact evaluation (see Section **6.4**). RCTs, in which people are randomly assigned to treatment and control groups, are the statistically strongest design for quantitative evaluation (see Section **5.1.1**), but quasi-experimental approaches are more common in quantitative evaluations than full RCTs. Statistically strong evaluation designs are rare and difficult to achieve in development settings (**Bamberger et al., 2011b, p. 20**) – and even more rare in humanitarian settings, where the constraints are correspondingly greater.

Although the term counterfactual is conventionally used for statistically valid counterfactuals, it is also possible to examine the counterfactual by using non-statistical quantitative and qualitative methods. Approaches as simple as asking the affected population what changes were associated with an intervention can help to establish the counterfactual, and triangulation can be used to support this approach. Such questions need to be carefully phrased to avoid provoking a positive response out of politeness. (See **Box 7** for a discussion on the use of recall.)

A paper from an American Evaluation Association meeting (**Bamberger et al., 2009**) lists a range of non-statistical ways of establishing the counterfactual – usually not 'what would have happened if the intervention had not taken place' but 'what would have prevailed if a given humanitarian action had not taken place'. The list includes a full range of qualitative and quantitative methods.

5.3 Desk review methods

5.3.1 *Key document research*

An initial literature search or desk study is the first step in any evaluation; it is an economic and efficient way of learning about the context, the humanitarian crisis, and the response before embarking on fieldwork. Key documents can often be found in the following ways:

- Ask key stakeholders and the reference group for suggestions.
- Check ReliefWeb (**http://reliefweb.int/**) and other key humanitarian sites such as AlertNet and the Humanitarian Practice Network.
- Look for bibliographies on the topic.
- Look in the ALNAP evaluative reports database for any previous evaluations.

- Search the Internet, including Google Scholar.
- When you find a useful document, check it for a reference list or other leads to additional helpful works.
- Circulate a provisional document list to the team and the reference group and ask them to suggest additional sources.

Once you have located key documents, team members need to read them. They can either note important issues in an evidence table (see Section 5.2.4), or prepare a short abstract that highlights the key points. Because not all documents are equally accurate, it's a good idea to have a system for reviewers to record their assessment of the document.

To avoid being overwhelmed with reading material, focus on documents that tell you something new, provide information on a key point, provide a useful analysis, or that you are likely to cite in the evaluation report. You may also need to include documents that your end users expect you to consult, even if you do not expect those documents to be helpful.

Using bibliographic software such as EndNote or Zotero to record the documents used in your study can save a lot of time later when organising your bibliography. Enter an abstract for each entry when you create it. This makes it easier to produce an annotated bibliography, which can be very helpful. Specific comments – for example on the relevance of a source to a particular evaluation or your opinion of its quality – should be recorded in a separate notes field rather than as part of the abstract.

If you conduct a large literature review, consider making an annotated bibliography publicly available to assist other research on humanitarian topics.

Use DropBox, Google Drive, or other remote collaboration software to share the main document set. This allows automatic updating of each team member's local computer copy when anyone adds a new file or changes an existing one.

Any good guide on research methods covers document research, although the treatment may be brief. See for example Thomas and Mohnan (2007). Texts that concentrate on key document research, such as Hart (1998), tend to be focused on the literature search needed for a doctoral thesis.

Good practice example 10

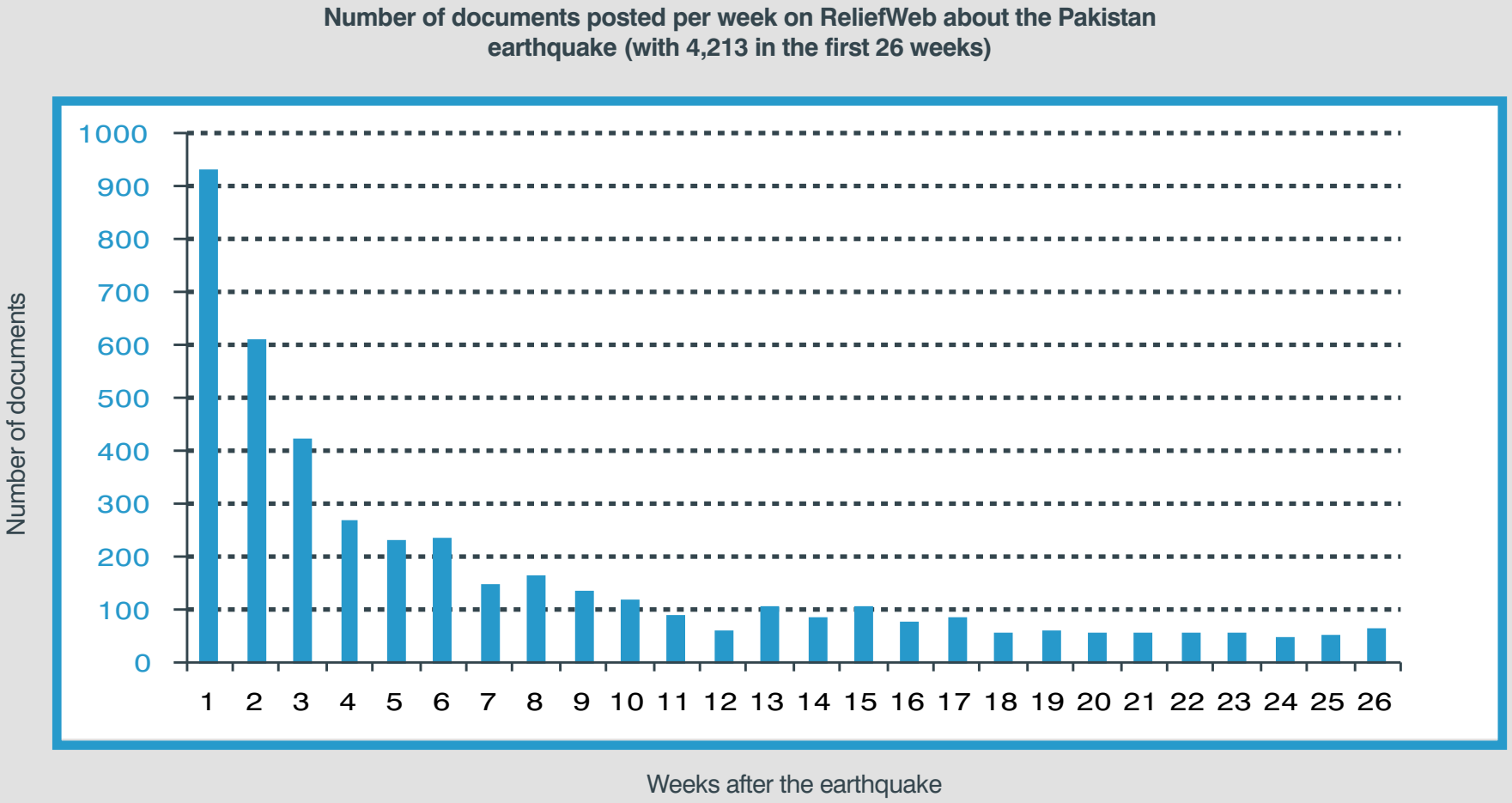
Publishing a bibliography of sources used for an evaluation

The 2009 follow-up to the Tsunami Evaluation Coalition's work published an annotated bibliography as a separate volume accompanying the main report. The bibliography contained 610 entries and was built in part on the bibliography developed for the Coalition's work in 2005 and 2006 (**Cosgrave et al., 2009a**).

5.3.2 Using large document sets

After recent humanitarian emergencies, agencies have generated a large number of documents describing their actions and reporting on progress. Even a single agency can produce many situation reports and other written materials in a short time. Many agencies post their reports on the ReliefWeb site (**http://reliefweb.int/**), run by the UN Office for the Coordination of Humanitarian Affairs (**Figure 14**).

Figure 14: Number of documents posted to ReliefWeb during a recent humanitarian response



Clearly it is not possible for the evaluation team to read thousands of documents. Many will be of only peripheral interest, but some will contain essential information. You can find the wheat in all this chaff by using software to index the documents in a way that allows you to do complex searches and measure keyword frequency.

Use a structured approach, such as a checklist or scoring tool, to abstract information in a rigorous way. When you have too many documents to use this approach, you can use a random sample.

5.3.3 Building a chronology

A chronology can be useful for understanding how the context and the response have changed over time and for highlighting linkages that merit further study.

Get the whole team involved in developing the chronology, as it can serve to familiarise members with the context.

Good practice example 11

Preparing a detailed chronology

The final evaluation of the Emergency Capacity Building Project presented a detailed chronology to explain the evolution of the project over time (Morris and Shaughnessy, 2007). The following extract from the chronology shows the first steps in the project.

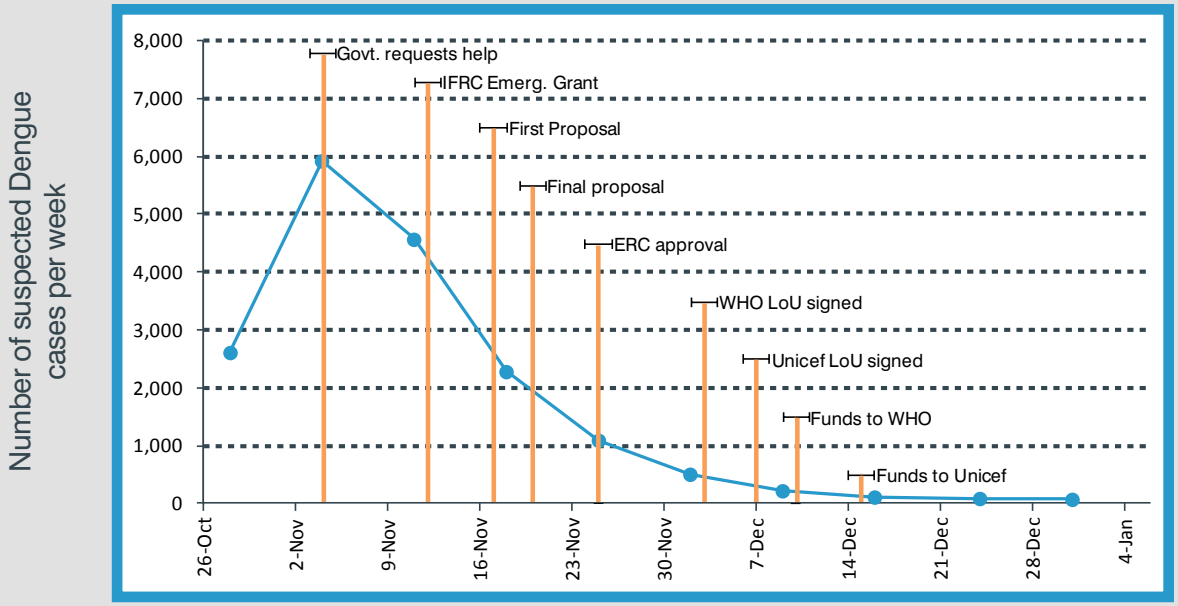
Date	Event	Description
March 2003	Creation of Inter-Agency working group (IWG)	The purpose of the group was to advanced more effective delivery of humanitarian assistance by non-governmental organisations (NGOs) through a joint capacity building strategy. The goals were to create sustainable new mechanisms to share best practices tap into the analysis of other specialist organisations, test innovative practices collectivrly in NGO field operations, and produce, share definitions and standards that can be applied collectively across the humanitarian sector.
July 2004	Report on Emergency Capacity	The review found that while, IWG members had made progress in building their emergency capacity, most notably in rapid funding and response systems, major capacity gaps remained, inhibiting both the speed and effectiveness of humanitarian action, as well as coordination with each other.
August 2004	ECB Project Grant Proposal	The ECB grant proposal, with CARE as the lead agency for seven agencies, was submitted to the Bill and Miranda Gates Foundation for assistance in developing initiatives 1, 2 and 3 of the ECB Project.

Chronologies are usually presented as a table of dates and events. Sometimes events that happen within an agency occur in a separate column from events taking places in the wider world.

Consider highlighting key points from a chronology in the form of a visual timeline (see **Figure 15**).

Figure 15: Key events in a 2009 dengue fever outbreak in Cape Verde

Timeline for the dengue outbreak in Cape Verde and the CERF application process



5.3.4 Content and keyword analysis

As discussed above, humanitarian responses can generate very large document sets. Over 100 documents a day were posted to ReliefWeb in the first week after the December 2004 Asian earthquake and tsunamis, and nearly 2,000 documents were posted the following month. It is not possible for an evaluation team to read all of these, but they can still be a useful source of information through content analysis.

Content analysis
Analysis of textual information in a standardised way that allows evaluators to make inferences about the information.

Content analysis typically takes the form of coding documents to reduce the text to a set of categories. One form of content analysis used in a number of humanitarian evaluations, including the recent CERF evaluation (Channel Research, 2011), is keyword analysis.

Keyword analysis
A form of content analysis that examines the frequency of occurrence of keywords to highlight trends over time in a single document set or to compare two document sets.

A keyword list might have up to 500 terms. Some categories will be represented by multiple keywords. For example, keywords for the water, sanitation, and hygiene category include the following:

- borehole
- chlorination
- faeces
- latrine
- sanitation
- spring
- wash
- water supply
- wells
- chlorine
- defecation
- hygiene
- night soil
- soap
- toilet
- water
- watsan

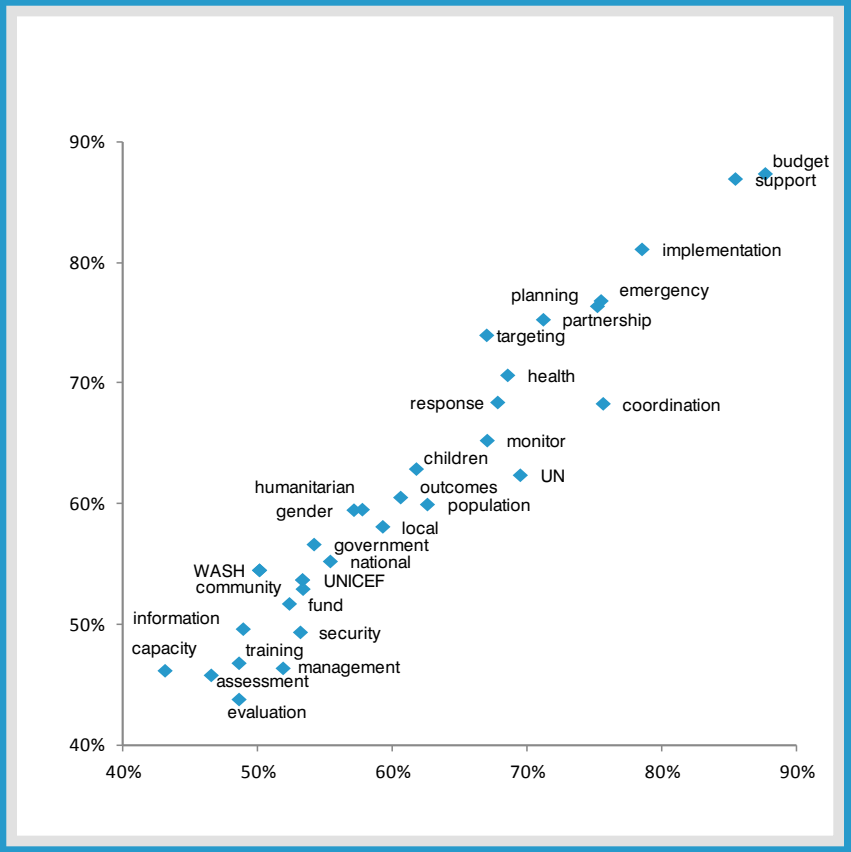
Two useful guides to textual analysis are Krippendorff (2004) and Benini (2009), a short guide to using three textual analysis tools for humanitarian and development workers.

Good practice example 12

Using keyword analysis to check that a sample is representative

In the CERF 5-year evaluation, countries were selected by the steering group for detailed studies based on a purposive sample rather than a random sample. The evaluation team wanted to make sure that there was no fundamental difference between the project portfolios of the 16 study countries and those of the 63 counties not selected. The prevalence and frequency of 405 keywords representing 327 categories in 7,797 documents were calculated using the dtSearch software package. The frequency of occurrence was compared between the 2,897 documents from study countries and 4,900 documents from the other counties to ensure that the projects in the sample were representative of whole population.

Comparision between two document sets for occurrence of a given term



5.3.5 *Numerical analysis*

Humanitarian operations can generate large amounts of numerical data, for example on funding, distribution, or number of people receiving assistance. Analysing these data can highlight particular issues. It can establish the following:

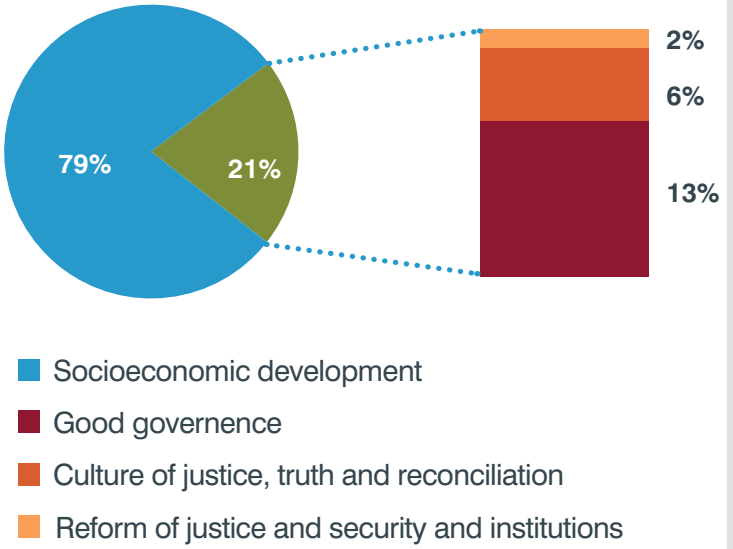
- Averages *and norms*
- Outliers (unusual cases)
- Whether patterns of activity were consistent with a project's stated objectives.

Good practice example 13

Analysing donor portfolios to see how funds were used

A multi-donor joint evaluation of support for conflict prevention and peace building in South Sudan included a portfolio analysis that examined spending on 2,724 projects and characterised the expenditures by peace-building category. As shown in the figure below, the analysis showed that nearly 80 per cent of donor support went for socioeconomic development (Bennett et al., 2010).

Distribution of budget of USD4.05 billion by Peacebuilding category



Spreadsheet software allows more flexible analysis than does database software. Learn how to use pivot tables and database functions in your spreadsheet software to do the analysis.

5.3.6 Cost-effectiveness and efficiency

Cost-effectiveness is a challenging aspect of humanitarian evaluations. It is being given increasing attention by donors in the drive to get the greatest possible impact for their funding. Normally EHA compares different approaches, because we cannot set a monetary value on human life and suffering. However, even comparing the costs of different approaches to the same effort can be difficult for two reasons:

1. There are not enough data to enable valid comparisons.
2. It is difficult to assess the trade-off between speed and cost.

One area in which data are unreliable is the estimation of unit costs. Typically, reporting from an agency to a donor includes the inputs and their costs. However, there is no standardised terminology for inputs, and similarly named inputs from two agencies may not actually be comparable.

For example, two agencies may provide latrines in the same camp for internally displaced people at very different unit costs. One agency may provide only the latrine slabs, to families in readily accessible areas, while the other trains community groups to manufacture the slabs instead of buying them, so that they can install latrines when they return to their villages; provides materials for building the latrine superstructure; works in relatively inaccessible areas; and also provides an extensive hygiene education programme.

Financial data are also often not usable for comparisons. Financial reporting to one donor does not include contributions from other donors, and financial reports often group a range of activities together under a single budget item. To extend the latrine example, one agency may use free transport provided by the military while the other pays for its transport; one may have only one donor for its latrine programme while the other has several.

Assistance delivered may also be difficult to establish. A 2011 joint evaluation of assistance to south and central Somalia was unable to determine precisely how much of the assistance was reaching the targeted population; estimates by interviewees ranged from 30 per cent to 70 per cent (Polastro et al., 2011a, p. 48).

When interviewing senior agency managers, ask them about the unit costs for different elements of their operation. Managers who are aware of their unit costs and cost drivers are more likely to be actively seeking to keep their costs down to maximise the assistance that they can deliver.

Despite the difficulties, it can still be useful to calculate unit costs, for three reasons:

1. It highlights qualitative differences between programmes that can be followed up during fieldwork.
2. It calls into question whether there is a need for all the services provided by agencies with higher costs or whether the affected community could easily provide some of the services itself.
3. It identifies approaches that deliver assistance at lower cost.

When evaluating cost-effectiveness, look for activities that have been carried out by more than one organisation and for which the number of inputs (latrines, medical consultations, people served) are clear. Try to estimate the unit costs for these, and if necessary ask the organisations for more detailed reporting on the costs.

When there are radical differences in unit costs, consider including examples of the different approaches to the projects to be sampled in the field, to see if the different cost levels have led to different outcomes and impacts.

A more difficult area of comparison is where a high-cost approach delivers assistance more quickly than a low-cost approach. Quicker responses may save lives and reduce suffering, but this is difficult to quantify. The classic example is airlifting relief supplies instead of transporting them by road. NGO-run nutrition programmes may have a lower unit cost than government programmes, but the latter are far more sustainable.

In such cases, evaluators should look at the whole chain, not just at the element that was handled quickly, to see if the more expensive option did in fact deliver faster assistance. For example, in some cases during the 1994 Rwanda relief effort, air operations continued after adequate land supply mechanisms were established, and goods that were readily available locally were airlifted in (Borton et al., 1996, chapter 5). It may be that an airlift takes so long to organise that trucks arrive before the first aircraft does, or that airlifted goods spend so long in customs or at the airport that there is no speed advantage.

It can be very instructive to examine the different unit costs for operations that are directly implemented and those that are implemented through partners. In some cases, one agency may operate both types of projects.

5.3.7 Assessing attention to gender in desk studies

Gender is a key issue in humanitarian action and a strong determinant of vulnerability (see **Box 5**). Assessing gender sensitivity is a challenge in desk studies. One approach is to classify the degree to which projects focus on gender, and to calculate what percentage of projects address the issue at different degrees of focus. The Inter-Agency Standing Committee has classified four levels of attention to gender (**Table 16**).

Contents

Table 16: The Inter-Agency Standing Committee's gender analysis tool

Level	Description
0	Gender is not addressed in any component of the project.
1	The project is designed to contribute in some limited way to gender equality. Gender dimensions are meaningfully included in only one or two of the three essential components: needs assessment, activities, and outcomes.
2a	Gender mainstreaming: The project is designed to contribute significantly to gender equality. The different needs of women/girls and men/boys have been analysed and integrated well in all three essential components: needs assessment, activities, and outcomes.
2b	Targeted actions: The principal purpose of the project is to advance gender equality. The entire project either (a) targets women or men, girls or boys who have special needs or suffer from discrimination or (b) focuses all activities on building gender specific services or more equal relations between women and men.

Source: (IASC, 2010a, 2010b).

Feedback

Contents
1
2
3
4
5
6
7

The same approach can also be applied to other cross-cutting issues such as vulnerability, age, HIV/AIDS status, and environmental sustainability to categorise projects as not focused on the issue, focusing on it in a limited way, addressing it in a significant way (but not as a principal focus), or treating it as the principal focus. HelpAge International (2010) used a similar categorisation approach in a recent study of humanitarian funding for older people.

If your study includes key information interviews, you should also ask questions about the issue in the interviews to triangulate the information from the portfolio analysis.

If there are too many projects in the portfolio to analyse all of them, you can select a sample. However, such a sample must be a random sample if you want to generalise from it to the whole portfolio.

5.3.8 Online surveys

These are much easier to organise than enumerator-administered surveys. They can be a useful adjunct to other data collection methods. However, they suffer from the potential bias involved in non-response. They are useful for identifying issues for deeper investigation by surveying the staff of an organisation or its partner organisations. The principal reference for such surveys is Dillman et al (2009). For a discussion of field surveys, see Section 5.4.5.

Good practice example 14

Using online surveys to gather views from a diffuse group

The Stay and Deliver review of good practice in delivering humanitarian assistance in different security and risk environments looked at a variety of strategies humanitarian workers have used to maintain an operational presence and continue their work. The team conducted fieldwork in six countries and did desk studies of another six. In order to capture the views of national staff, the team developed an online survey in English, French, Spanish, and Arabic to gather the views of 1,148 local staff (Egeland et al., 2011).

Always ask respondents if they would be willing to answer further questions or to provide clarification by phone or email.

Consider translating the survey into other languages to make it accessible to everyone you want to survey.

5.4 Field methods

5.4.1 Key informant interviews

Key informant interviews are the backbone of humanitarian evaluations; often, aspects of a humanitarian response are not well documented, and asking those involved is the best route to establishing what happened and why. Key informants are people who can be expected, based on their role (for example, as agency staff or community leaders), to know about the intervention being evaluated. Key informants who have reflected on the operation and drawn some general conclusions are particularly useful sources.

Key informant interviews are an efficient way of gathering information when time and resources are limited. Interviews typically take 45 minutes to an hour and may involve one or more interviewees. The interviews are semi-structured, using an interview guide or checklist.

Developing an interview guide is an art, as the wrong question may generate an automatic 'correct' answer rather than a considered one. For example, the ToR may include the question 'to what extent has the project considered gender in the design?' – but that phrasing might simply evoke a repetition of the agency's current policy on gender. If instead you ask 'who benefited most from this project – men, women, boys, or girls?' and follow up on the response with further questions, you are far more likely to get an indication of the extent to which project managers are aware of the importance of disaggregating data, or have done a gender analysis of the intervention. Remember, it is the evaluator's job, and not the interviewee's, to answer the questions in the ToR.

When planning your interview guide, consider the following:

- Limit your questions to 20–25, the most that you can get through in an hour-long interview.
- Begin each interview with an explanation of the interview's purpose and the interviewee's rights, including confidentiality.
- Sequence your questions so that initial questions build rapport and more contentious questions are asked towards the end. In a semi-structured interview, you can alter the order of the questions to maintain a natural flow. Be sure to follow up on any unexpected but relevant information.
- Always include a final question about what the interviewee has learned or would do differently the next time.
- Asking if there is anything else that an interviewee was expecting you to ask can highlight overlooked areas, help resolve misunderstanding about the purpose of the evaluation, and sometimes provide new leads.
- Always ask the interviewee to suggest other people to talk to.

Depending on the culture, some questions may be regarded as sensitive, and the answers may be dictated by social norms. For example, if you ask beneficiaries, 'how useful was the assistance that you received?' they may feel obliged to praise the assistance out of politeness. If instead you ask them, 'how do you think your neighbours viewed the assistance that they got?' they may be less constrained by good manners.

Box 7: Using recall

Sometimes there are few historical or baseline data to indicate what things were like before the crisis. You can help compensate for this by asking interviewees how things are now compared to the past. People sometimes remember the past more favourably than it really was, so it is important to cross-check by asking further questions. For example, if you are told in an interview that every family had an oxcart before the big drought 10 years ago, ask what sort of cart the interviewee or family members had. Asking about costs and transport charges can help establish the level of accuracy.

Recall is not a substitute for a baseline study. It becomes less accurate over time, especially when the recalled fact is not of central importance to the interviewee (de Nicola and Giné, 2011). The accuracy of recall also varies with the social situation of the interviewee (Manesh et al., 2008).

Contents
1
2
3
4
5
6
7

If you use recall, restrict it to important events rather than background data, and try to introduce an element that can be verified from other sources.

Critics of semi-structured interviews argue that they collect perceptions, not facts, and that interviewees may give biased answers or answers that fit their own agenda. These criticisms are true to an extent. However, no research method is totally free of bias. Even in the physical sciences, the observer influences the observation, hence the use of double-blind trials. Even in quantitative methods like surveying, the response to a survey question depends on who is asking it (Davis et al., 2010).

Perceptions are valuable information in their own right; for example, they are used as the basis of some key economic indicators such as the Business Confidence Index and Purchasing Managers Index. Like these indices, semi-structured interviews capture the learning and perceptions of people with relevant experience.

Clearly, all interviewees have an agenda. Some may want their work to be widely known, others may want to conceal flaws, and yet others may want to see assistance continued. The evaluator needs to remain aware of the potential for such biases and aware that interviews are constructed processes. When using interview data, consider what weight to attach to different views.

- Give greater weight to the views of those most affected. For example, the views of beneficiaries as to whether they were consulted or not have far greater weight than the views of agency staff on the same topic.
- The views of someone with significant experience in a sector might have greater weight than those of someone with little experience. However, this is not always the case – a person with a lot of experience may be locked into a particular approach or mind-set.
- Consider giving greater weight to views that contradict the apparent self-interest of the interviewee. For example, pay close attention when beneficiaries argue for curtailing assistance or government staff call for channelling resources through NGOs.

Ideally, evaluators should be able to speak the language of the key informants they interview. When this is not possible, they need to work with interpreters, which increases the time required to work through a set of questions. Evaluators should make every effort to engage good quality interpreters. Professional interpreters are well worth their fee as they are less likely to distort the question or responses. In one recent evaluation in Afghanistan, the community elders were asked about the impact of the Taliban on the government services that they got. The agency staff person acting as interpreter told the team that the elders complained that the actions of the Taliban had reduced their access to services. Later, the evaluation team's assistant told them that the elders actually said that they were the Taliban and that the government was refusing to provide them with services unless they surrendered their weapons.

If you have to rely on agency staff to interpret and you have no knowledge of the language, it is good to have an assistant who understands the language and can brief the team if there are any major inaccuracies. You can also openly record exchanges for later analysis to encourage accurate interpretation.

5.4.2 Interviews with the affected population

Interviews with the general population are a rich source of data for humanitarian evaluation and an essential step in meeting the needs of the affected population and consulting them. They often provide very good information on the assistance delivered, especially its appropriateness.

DESIGNING YOUR APPROACH

Interviews with affected populations have special challenges (**Table 17**). Local knowledge is essential in order to be aware of these challenges and address them ethically.

Contents

Table 17: Challenges and ethics of consulting the affected population in a humanitarian crisis

Challenge	Ethical response
Population groups are at risk of abuse or targeted violence.	If talking to an evaluation team may increase people's risk of abuse or violence, do not interview them unless it is safe to do so.
Politically powerful groups dominate the consultation process.	Be alert to power dynamics in the affected population, and seek ways of reaching marginalised and less powerful people.
Members of the affected population are traumatised by their experience of the crisis.	Interview sensitively to avoid distressing interviewees, for example by causing them to relive a traumatic experience.
There has been a breakdown of trust and relationships within the affected population as a result of the crisis.	Ask trusted members of the community to introduce the evaluation and consultation process. Ensure complete transparency in explaining the purpose, constraints, and intended use of the evaluation, and in explaining how data will be used and stored.

In most environments, unless the interview is in a private space such as the home, a one-on-one interview will quickly turn into a group interview as friends and neighbours join in.

Individual beneficiary interviews are more likely to provide an accurate picture of what assistance has been received, whereas group interviews may evoke a more 'political' stance as interviewees play to their audience as well as answering your question.

Steps to follow in designing a process of consultation with the affected population include the following:

- Identify which population groups you want to interview during the evaluation, and why. Remember that local knowledge is key to good decisions.
- Should you include direct beneficiaries of the humanitarian action as well as those who have not benefited but were affected by the crisis?
- How should you disaggregate the population groups to be interviewed – for example, by gender, age, ethnicity, socio-economic status, and status as displaced person, refugee, or resident?
- Choose an interview approach (see **Table 18**).
- Be prepared to be flexible about your consultation methods if lack of time, lack of access, or logistical constraints require a change of plan.
- Be sensitive to the particular challenges and ethics of consulting a disaster-affected population (see **Table 17**).
- Describe your interview process fully in the methodology section of your evaluation report.

Good practice example 15

Consulting with the affected population

In a UNHCR evaluation of age, gender, and diversity mainstreaming (**Thomas and Beck, 2010**), the country study for Colombia used a highly participatory approach. Four communities in different parts of the country were selected to participate in the evaluation through the following steps (**Mendoza and Thomas, 2009**):

- A workshop with community members to construct a timeline of events leading up to UNCHR's participatory assessment (a key tool in the mainstreaming process) and the resulting action plan – to gauge awareness of the action plan and to evaluate whether the participatory assessment had generated changes in the community
- Meetings with sub-groups of women, men, adolescent girls and boys, children, older people, and people with disabilities – to ascertain whether different groups had different perceptions of the results of the participatory assessment and action plan

AN OVERVIEW OF DIFFERENT METHODS

Most interviews with the affected population take the form of group interviews. These are often incorrectly called focus group interviews. Focus groups are intensive facilitated discussions with six to eight people from a particular population group with a shared experience (see Krueger and Casey, 2009). In practice, it is rarely possible to control the environment enough to maintain a focus group; most quickly evolve into a less controlled group interview. Ideally, a group interview should involve no more than 10 or 12 people, but the number can rapidly increase to as many as 100 or more. **Table 18** provides an overview of the different ways of consulting the affected population.

Table 18: Ways of consulting the affected population

Method	Advantages	Disadvantages
Focus groups and group interviews (6 to 10 people)	This is an effective way of seeking the views of a homogenous group – for example, internally displaced people or a particular livelihood group – which may be further disaggregated into men and women. It is suited to open-ended discussion, which may be guided by a checklist for consistency but allows interviewers the flexibility to follow up on interesting information – for example, by asking 'is that the same for everyone?' or 'do the women agree with that?' – and to put questions directly to individuals. Interviewers can use participatory rapid appraisal methods with the group (see Section 5.4.4).	It may be hard to control the size and composition of the group; if others join spontaneously, it may affect the dynamic of the group. This approach requires a skilled facilitator and a note-taker. It may be difficult to ask sensitive questions in a group.
Community meetings	It is possible to use existing community structures to consult the affected population, although this choice needs to take local power dynamics into account. It is possible to reach a large number of people, possibly even everyone in the community.	Meetings may be dominated by community leaders. Women and young people may be reluctant to talk openly. It may not be possible to disaggregate responses sufficiently to enable useful analysis.

Method	Advantages	Disadvantages
Individual and household interviews	It may be easier to ask about sensitive issues. Interviews can yield more in-depth and precise information. Interviews could be used for illustrative household case studies. This approach can usefully complement group interviews.	This approach is more time consuming because it requires more interviews. A smaller sample may not support generalisations, although it could still be illustrative.
Semi-structured key informant interviews	These can provide a good overview of the programme. They may provide insights that can be followed up in focus group discussions or individual interviews. Following a checklist ensures consistency but allows flexibility to follow up on interesting information.	Good local knowledge is necessary to select appropriate key informants and to be aware of their potential biases.
Formal surveys	These provide useful statistics and comparable data sets. The resulting data make it easy to generalise as long as the sampling method is sound. Quantified data may be more convincing to decision-makers.	They may be too time consuming and expensive to use when there are limited resources – for example, they require careful design and testing before implementation. They may not be feasible in a conflict setting or other unpredictable environment, where access is constrained, or where people are distrustful and reluctant to respond. They allow little leeway for following up on interesting or unexpected responses.

One tried and tested approach is contracting out detailed beneficiary interviews to a social researcher. The team can identify key issues from this research to investigate further in their own beneficiary interviews in the field.

5.4.3 Observation

Direct observation is very useful, especially for triangulating data gained from other sources. Observation is also useful for gathering primary data, but may need to be supported by other methods, as the meaning of what you see is not always obvious.

Carry a small camera for capturing visual observations where it is culturally acceptable. Photographs are a powerful way of showing staff at headquarters what the issues are in the field.

Good practice example 16

Using photographs to support key messages

In the response to the Kosovo crisis, an agency with substantial experience in the water and sanitation sector carried out a number of projects that fell below their usual standards. During the introduction to the debriefing meeting for the agency, the suggestion that the agency had failed to meet its own standards was strongly rejected by agency staff. However, the atmosphere changed when the evaluator used a number of photographs, like the one shown here of a precariously balanced 5-tonne water bladder, which was located in an environment full of children, to illustrate neglect of safety rules and agency policies (**Wiles et al., 2000**).

If possible, have a look around before engaging in individual or group interviews with the affected population. The evaluator is usually seen as part of the external aid machine and as having the potential to influence the flow of aid. In some cultures, this may colour responses to questions. Looking around first may provide you with information that allows you to challenge responses and promote a franker exchange.

When evaluators expect to see a number of similar cases, it can be very useful to structure the observations so that it is easier to identify similarities and dissimilarities between the cases. This can take the form of a simple checklist.

Evaluators should always be on the lookout for unobtrusive measures that provide an indirect indicator of key data. A well-known example is the wear on carpet tiles in a museum indicating which exhibits were the most popular (Webb et al., 1966). Evaluators may use the absence of queues at latrines (or the absence of uncontrolled defecation) as an indicator of the adequacy of latrine provision. In one evaluation in Malawi, the lack of the traditional sung greeting on arrival in a village (in contrast to the tuneful welcome at other sites) suggested to the evaluator that the villagers were not very happy with the work of the agency being evaluated.

5.4.4 Participatory rapid appraisal methods

Participatory rapid appraisal
A set of methods enabling local people to enhance, analyse, and share their knowledge and learning in ways that outside researchers can readily understand.

Many participatory rapid appraisal (PRA) techniques are visual and can capture a large amount of qualitative information in a diagram, map, or picture. Others are useful for relative measurements or values. Most are accessible for both literate and illiterate participants. Thus, these techniques are a way in which groups and communities can carry out their own analysis, in a facilitated way, and in a format that external researchers can understand. These activities can usefully be part of a group interview. **Table 19** provides examples of PRA techniques that can be used in EHA.

Table 19: Participatory rapid appraisal techniques

Technique	Description	Potential use
Calendar	The group constructs a diagram that shows changes over time – for example, in agricultural work, gender-specific workloads, or disease.	A 24-hour calendar for women could show the amount of time spent accessing relief resources (such as food aid) in the early stage of a crisis. A seasonal calendar could show periods of greatest food scarcity, which could then be compared in the evaluation with the timing of food distribution.
Timeline	The group constructs a timeline of events.	One timeline could record key events or moments of insecurity during a conflict-related crisis, and a second could record when humanitarian assistance was provided.
Proportional piling	The group is given a pile of 100 stones, beans, or seeds to represent a total in a given category (such as household income) and is asked to divide the pile up to illustrate the relative significance of various elements within that category (such as sources of income)	Knowing the relative significance of different sources of livelihood can indicate the relative significance of a specific relief intervention, such as a cash-for-work scheme or food aid. This technique can also be used to identify the poorest in the community.
Ranking	The group is asked to rank different items, either against each other or two-dimensionally according to certain criteria.	Ranking could be used to understand how well different types of humanitarian assistance (such as food aid, NFIs, seeds, and tools) meet recipients' needs. If this is done two-dimensionally, it could capture different needs in the household – for example, of elderly people, women, men, or children.

Technique	Description	Potential use
Transect walk	The evaluator walks through the village with a small group or with key informants and asks about whatever he or she observes – for example, who has access to grazing in a pasture or who lives in a home.	This can help evaluators understand the different impacts of a natural disaster on different areas or groups in a community and to explore whether the humanitarian response was sensitive to these differences.
Venn diagram	Circles representing different categories overlap where the categories hold a given value in common.	These diagrams are usually used to show different institutions – for example, international, national, and local humanitarian organisations – how they relate to each other, and their relative significance in responding to a crisis.

You can find further information on the use of PRA tools in *Methods for Community Participation* (Kumar, 2002) and *Participatory Rapid Appraisal for Community Development* (Theis and Grady, 1991). The PRA approach stems from the work of Robert Chambers Chambers (1994). Online resources on the tools have been provided by the World Bank (Bank, 2007) and the Canadian International Development Agency (CIDA, 2005).

PRA techniques look simple but require good facilitation skills. They should be used with a relatively homogenous group, such as recently displaced women or residents who remained in a village after it was attacked. It can be useful to carry out PRA exercises with men and women separately.

Good practice example 17

Using participatory rapid appraisal tools

The 2001 Disasters Emergency Committee evaluation of the response to the 2001 Gujarat earthquake used PRA tools for community research. First, key research topics were developed from the evaluation criteria in a workshop with the participation of members of the research team. Three participatory exercises – a matching game, a timeline, and a ranking exercise – were then developed to facilitate discussions with the community groups on each topic.

For the matching game, each group was asked to draw, or write onto cards, the types of aid that the community as a whole received following the earthquake. The cards were laid in a line on the ground. The group then prepared a second line of cards showing who had received the different interventions. Group members were then given cards representing timing (too late, too early, on time), quantity (too little, too much, OK), and quality (too low, too high, OK) and asked to place them below each line.

For the timeline, the group was asked to mark interventions, and the extent of the community's involvement in them, on a timeline of the earthquake response, and to comment about where in the process they would like to have been involved.

For the ranking exercise, the group was asked to list all local capacities that were important at the time of the earthquake – including organisations, structures, external contacts, networks, skills, resources, knowledge, and key people – and then to rank with stones the strength of each capacity before, during, and after the earthquake.

After each exercise, the evaluators discussed the results and the role of the evaluated agencies in any change (**Humanitarian Initiatives UK et al., 2001**).

5.4.5 Field surveys

Surveys can be used for a wide range of purposes, from assessing malnutrition to assessing intent to return (refugees) or the effectiveness of health education messages. **Fowler (2009)** provides a good, succinct introduction to surveys. This section deals mainly with enumerator-administered structured face-to-face surveys. Section **5.3.8** deals with online surveys. There are also a number of specialist surveying techniques (for example, for nutrition) that are not addressed here. Computer-assisted telephone interview surveys are not generally used in humanitarian evaluation.

Surveys can provide useful data on beneficiary perceptions of aid. Always consider doing a beneficiary survey if you want to know what overall perceptions of the assistance are.

Surveys are complex and need careful management. They are one of the main quantitative evaluation methods. Their strength stems from the fact that if a sample is randomly drawn from a given population, it is likely to be an accurate reflection of the whole population.

Structured face-to-face surveys are seldom used in humanitarian evaluation for a number of reasons:

- **Cost** – a professionally administered survey can cost between US$20,000 and US$100,000. Such surveys are most often found in well-resourced joint humanitarian evaluations such as the Tsunami Evaluation Coalition evaluations (Telford and Cosgrave, 2007) or the 2010 follow-up evaluation (Brusset et al., 2009).
- **Time** – developing and piloting a survey instrument, training the enumerators, conducting the survey, and processing the results, takes longer than other evaluation methods[8].
- **Lack of a good sampling frame** (a list that approximates the actual population) from which to draw a random sample – surveys in a humanitarian context sometimes use pseudorandom sampling to get around this.
- **Lack of baseline surveys** – surveys are most useful when there is a baseline to compare the results with. This is rarely the case in humanitarian evaluation.
- **Lack of evaluators with the requisite skills** – most humanitarian evaluators come from a qualitative social science background and don't have much experience of formal survey work.

[8] The survey is often the slowest data collection method. The first draft of the Disasters Emergency Committee evaluation of the response to the tsunami, which was based on qualitative methods, had to be revised when the data from the quantitative beneficiary survey became available (Tony Vaux, personal communication).

The next step is to design your survey instrument. (Normally you would start with this, but given the difficulties of finding suitable sampling frames in humanitarian contexts, it is better to start with the sampling frame, as this might influence the survey approach.) Survey questions, though simpler in some respects than semi-structured interview questions, are more complex in other respects, especially as the interviewer cannot adjust the question and probe for follow-up responses as is possible with semi-structured interviews. Structured face-to-face surveys are based on the assumption that the questions will be asked in the same way of each interviewee. Survey questions need to be carefully designed and then tested before being used, as they cannot be adjusted once the survey is underway.

The usual way to test survey questions is to conduct a pilot survey. How questions are asked can have a significant impact on responses, and a pilot survey may highlight some of these issues. **Fowler's (1995)** monograph on survey questions is a useful resource.

If the survey has to be translated, have the translation translated back into the original language by a different translator.

The number of enumerators needed depends on the sample size and distribution, as well as the sampling strategy. The sample size you need depends on a number of factors including the following:

- How common your study topic is in the population – for example, if you are surveying the intention to return, and only 5 per cent of people are interested in returning, you will need a far bigger sample size to draw any statistically valid conclusions than for a situation in which 75 per cent of the population intend to return.
- What level of error you are prepared to accept – the smaller the acceptable level of error, the larger the sample size must be. The most common designs have a 5 per cent possibility of a false positive (such as finding that the assistance was effective when it actually was not) and a 20 per cent possibility of a false negative (such as finding that the assistance was not effective when it actually was).

Always pilot your survey instrument before use. Consider recruiting a local market research company to conduct the survey. Some agencies may conduct surveys for their own purposes that could be of use to the evaluation.

Box 8: Taking sex and age into account during field data collection

The following steps will help ensure that your evaluation is gender- and age-sensitive and yields data that can be disaggregated by sex and age:

- Make your fieldwork teams well balanced between men and women. Female interviewers and translators usually have the best access to women and girls in the affected population.
- Ensure that your key informants include women who are knowledgeable about the community and about the particular needs of women and whether and how they have been met. These could include female teachers, leading market women, and female nurses. Ensure that some key informants are also knowledgeable about the needs of children, youth, and older people.
- Ensure that at least half of the participants in focus groups and group interviews are women. Ideally, women should be interviewed separately from men, who tend to dominate mixed-sex discussions.
- Hold age-set focus groups and group interviews, for example with children and youth and old people. Adults tend to dominate mixed-age discussions.
- Keep a record of the gender of key informants and other interviewees so that you are aware of any potential gender bias in the information you have gathered from interviews.

Source: Based in part on Mazurana et al (2011).

After you have collected your survey data, you need to enter the data and clean it. This can be expensive and time consuming, especially if the survey contains open-ended questions.

Using electronic aids for collecting survey responses can almost eliminate the cost of data entry and cleaning, as the survey form can be set up to reject invalid combinations in responses, and the collected data can simply be uploaded to the main database without any further entry. This can also speed up the survey process.

Digital tools can be used for assessment surveys as well as for evaluation surveys. The Multi-cluster Rapid Assessment Mechanism in Pakistan used personal digital assistants (PDAs) to gather assessment data (McRAM Team, 2009). Shirima et al. offer an example of the use of PDAs in a survey of 21,000 households in Tanzania and provide some insight on the practical issues involved (Shirima et al., 2007). Recent research suggests that using PDAs has the potential to reduce the logistics burden, cost, and error rate of data collection (Seebregts et al., 2009).

A new development is the growing use of mobile phones for data collection. These have the advantage that data are uploaded immediately, preventing the risk of data loss if the phone is damaged or lost. This also allows for real-time quality control and data-collector supervision to reduce the risks of data fabrication (Tomlinson et al., 2009). Mobile phones offer similar advantages to PDAs, in terms of reducing recording and entry errors (Zhang et al., 2012). Several software packages are now available for setting up mobile-phone-based surveys.

5.5 Learning-oriented approaches

Learning-oriented evaluations are usually more participatory and should involve those who need or want to do the learning. Methods should be conducive to reflection and the admission of mistakes as well as successes (see Section 3.1). This section presents a number of tried and tested approaches to learning-oriented EHA.

In the activist culture of humanitarian operations, it is often said that there is not enough time for participatory and reflective exercises. But only a small portion of humanitarian action is so time-critical that it would be harmed by taking time out for evaluation. Planned and facilitated well, a number of learning-oriented activities can be completed in a couple of hours, or a day, and the value of pausing to reflect and learn collectively merits the time investment.

When designing a learning-oriented evaluation, first identify who is the focus of the learning – for example, specific staff members, partners, or members of the affected population. Then design processes that enable them to reflect.

These include the following:

- Create a safe space for reflection and discussion where participants feel they can talk freely and admit mistakes.
- Ask simple, open questions, usually the most effective way of encouraging reflection (see 'Questioning and Listening' in Chapman et al., (2005)).
- Use participatory and creative processes.
- Ensure that the lessons learned are well documented.
- Provide good facilitation.
- Encourage those who do the reflection and learning to take ownership of the evaluation findings.
- Build in a process for follow-up.

5.5.1 *After-action reviews*

After-action review
A structured discussion of a humanitarian intervention that enables a team to consider and reflect on what happened, why it happened, and how to sustain strengths and improve on weaknesses.

An after-action review is normally a facilitated process, involving all those who have been identified as the focus for learning, in a workshop setting. An open atmosphere that fosters trust between participants is essential for a successful after-action review. The general principle is 'no attribution, no retribution'. A neutral and objective facilitator is essential to ensure that the discussion stays focused on issues, remains positive, and does not deteriorate into self-justification or blame. The review should address the following questions:

- What was the objective or intent of the action?
- What actually happened?
- What went well?
- What could have gone better?
- What would we do differently next time?

An after-action review includes the following steps. For each step, individual participants could write their responses on sticky notes, with a facilitator clustering the notes to reflect particular themes, or the group could respond to each question collectively, with the facilitator writing the responses on a flip chart.

1. Participants, individually or in pairs, write down their understanding of the objective or intent of the action. (Because humanitarian action often takes place in chaotic environments and plans can quickly become out-dated, this step is a useful way of finding out if the objective is clear and shared, and redefining it if necessary.)
2. Participants then write down what actually happened – possibly working as a group to construct a timeline of key events and changes over time in the situation or the programme.
3. Next, the group addresses three questions: What went well? What could have gone better? What would we do differently next time?

4. This is followed by a discussion, at the end of which the facilitator summarises all the lessons that have emerged and asks participants to vote for what they regard as the three most important lessons.
5. The key lessons learned, and any actionable recommendations that emerge, are documented and circulated to all participants. A time-frame, for example 6 or 12 months, may be agreed on for assessing progress towards implementing the recommendations.

Several guides to conducting after-actions reviews are available, including the **USAID guide**, the Food Security Information for Action's After Action Reviews and Retrospects.

Field staff often appreciate the opportunity that an after-action review gives them to reflect and to consolidate their learning.

5.5.2 Storytelling

Telling stories is a natural way for most of us to communicate. Focusing and facilitating this process can be a creative way to facilitate learning, especially when it is accompanied by a process of questioning and reflection to deepen analysis. This section describes two ways of using storytelling for learning purposes.

USING METAPHORS

This is best facilitated in a workshop, for example through the following steps:

Participants, individually or in groups, draw images (for example of a tree or a river) to represent their experience of the programme being evaluated. The different parts of the tree or river represent different parts of the programme.

1. Participants tell a story about the drawing, without interruptions, to their colleagues.
2. Those listening to the story ask questions about it to deepen analysis and learning and to offer their reflections.

3. At the end of this process, the group lists the main achievements, challenges, and lessons learned from the storytelling about their combined experience of the programme.
4. This can be focused on the future by asking the following questions at the end:
   a. *How can we build on what has gone well?*
   b. *How can we deal with the challenges?*
   c. *How will we apply what we learned here?*

Good practice example 18

Using metaphors and storytelling in a review

The Tearfund review of capacity building in disaster management used the metaphor of the tree as the basis for storytelling. In a monitoring and learning workshop, each Tearfund partner was asked to draw a tree.

- The roots represented values and principles underpinning the programme.
- The trunk represented partners (organisations and individuals).
- The branches represented activities.
- The fruit represented programme achievements.
- The leaves represented lessons learned.
- Broken branches on the ground represented internal challenges.
- Clouds in the sky represented external challenges.
- Buds represented activities planned but not yet implemented.

The storytelling was followed by a learning review. Despite some initial scepticism, participants said that they found this to be a visual and creative participatory process that had encouraged reflection and learning; led to the joint realization of achievements, challenges, and lessons learned; and provided a snapshot of the whole programme. They also recognised the challenge of translating learning into action.

THE MOST-SIGNIFICANT-CHANGE TECHNIQUE

The most-significant-change technique is a participatory approach that can be used with staff or local community members. When used with the affected population it can have an outcome or impact focus. It builds from the field level upwards, using stories that capture what happened, when, why it happened, and why it was important. This technique was used in the IFRC evaluation of the response to the 2007 Peruvian earthquake; participants were asked to identify the most significant changes that has occurred as a result of the intervention by the Peruvian Red Cross (Martinez, 2009).

The steps are as follows:

1. Stories are collected at the field level in response to two questions: During the last month [or other time period], in your opinion, what was the most significant change that took place in the lives of people participating in the project? Why was this the most significant change?
2. The most significant of these stories are selected by a panel of stakeholders or staff.
3. Once the changes have been recorded, the stories are read aloud, often in a workshop setting, and the participants discuss and reflect upon the value of the reported changes.

Advice for using this technique is offered by Davies and Dart (2005).

5.5.3 *Appreciative inquiry*

Appreciative inquiry
A method that seeks to renew, develop, and build on what went well in an intervention, based on the assumption that questions tend to focus attention in a particular direction.

Appreciative inquiry focuses on solutions rather than problems. It focuses on what works, based on the assumption that questions about strengths and successes are more engaging, illustrate possibilities, and can even be transformational. As a participatory process it engages key stakeholders and can help them develop their evaluative thinking and sense of ownership of the evaluation. When there is fear and negativity about evaluations, this approach can be energizing, build trust between the evaluators and those whose work is being evaluated, and engage reluctant participants.

In the planning stage of an evaluation, appreciative inquiry can be used to construct the logic model of the programme and to identify the evaluation questions through a participatory process, usually a facilitated workshop in which participants are guided through interviewing each other and agreeing on key evaluation questions as a group, in the following steps:

1. Ask participants about best practices, examples of success, and the values underpinning their work and the programme. This can be done by participants interviewing each other.
2. Ask participants to create a vision for the programme's future that is grounded in their past successful experiences.
3. Identify evaluation questions by asking participants: What questions should the evaluation address, so that you have the information needed to help you move towards your goals?

In the data collection stage of an evaluation, appreciative inquiry can be used to explore examples of success and what can be learned from them, as well as values and wishes. Questions about vision can explore how the programme could move to a higher level of performance. This will contribute the most to learning if it is carried out in a group interview format.

Preskill and Catsambas (2006) discuss appreciative inquiry in detail and provide examples of its use in evaluations in different sectors.

Considering using appreciative inquiry with people who may be sensitive about criticism.

Can this section be improved?

NOTES

NOTES

Please share your feedback:

Help us make the Guide more practical and user-oriented by sharing your feedback on its content and navigation. Feedback can include questions, suggestions, useful resources or practical advice from your own experience. You can share your comments and ideas by clicking on the feedback button at the bottom of any page. This will generate a confidential email to the EHA team in the ALNAP Secretariat. For more information on how you, your team or organisation can engage more actively in piloting the Guide contact us directly at **eha@alnap.org**

REMOTE, JOINT, REAL-TIME, AND IMPACT EVALUATIONS

Contents of this section

6.1 Evaluation when access is constrained **186**

6.2 Joint evaluations **188**

6.3 Real-time evaluations **192**

6.4 Impact evaluations **198**

How to use this section

This section explains several special types of evaluations used in EHA – evaluations when access is constrained (Section **6.1**), joint evaluations (Section **6.2**), real-time evaluations (Section **6.3**), and impact evaluations (Section **6.4**). These cutting-edge EHAs deserve further study in order to establish the best approaches and methods.

Tables, figures and boxes

Table 20: Alternative methods to use in situations of restricted access **187**

Table 21: Characteristics of a real-time evaluation **193**

Figure 16: Forward- and backward-looking aspects of real-time evaluations **195**

Table 22: Methods used in real-time evaluations **197**

6.1 Evaluation when access is constrained

Most EHA fieldwork methods require evaluators to be present in the field to observe and conduct interviews and exercises. This can be a problem when security issues prevents access by regional or international evaluators.

Consider whether certain evaluators might be less at risk, due for example to their nationality or ethnicity. Keep in mind, however, that you still have a duty of care towards all evaluators.

Security or access constraints affect the methods that an evaluation team can use. The following list of methods is ranked in order of increasing difficulty in situations of constrained access.

1. Document research
2. Numerical data analysis
3. Key informant interviews at headquarters
4. Key informant interviews in the field
5. Beneficiary interviews
6. Observation
7. Beneficiary focus group discussions
8. Beneficiary surveys
9. Participatory rapid appraisal

Local consultants may seem to be an option in situations where it is safer for them to work than it is for international evaluators. However, in chronic complex emergencies, everyone, including local consultants, may be perceived by the community as biased or associated with one faction or another. International consultants may be at risk because they are associated with parties to internationalised conflicts.

When putting the evaluation team together, consider how national consultants or international consultants from different countries may be perceived.

Table 20 lists alternatives to sending evaluators into the field.

Contents
1
2
3
4
5
6
7

Table 20: Alternative methods to use in situations of restricted access

Method	Alternatives
Key informant interviews in the field	Interviews by satellite phone Interviews at an external event attended by the informants
Beneficiary interviews and group discussions	Interviews and discussions with groups brought to accessible areas for training or other reasons, or who regularly travel to accessible areas (such as more secure market towns), or at the point of entry to safe areas
Observations	Satellite imagery, videos, and photographs made using cameras with built-in Global Positioning System data capture
Beneficiary surveys	Text messaging surveys and online surveys (these are subject to self-selection bias and need to be interpreted with care)
Participatory rapid appraisal tools	Use of the tools through a community dissemination process (the timescale for this normally exceeds the timescale for humanitarian evaluations)

Security issues that limit access for evaluations may also limit monitoring of projects, leaving the evaluation team with little reliable monitoring data. The IASC evaluation of south and central Somalia identified 11 lessons for good quality monitoring in such circumstances (Polastro et al., 2011a, p. 28). Triangulation – particularly of data from within the inaccessible area with data from outside it – is critical.

6.2 Joint evaluations

Joint evaluations are becoming more common wtarian aid sector, reflecting a wider trend towards 'jointness' in the aid world[9]. Joint evaluation can be found in multi-donor funding channels, for example the evaluation of support for peace-building in South Sudan (Bennett et al., 2010); in NGO alliances, for example the multi-NGO evaluation of the Yogyakarta earthquake response (Wilson et al., 2007) and the ACT evaluation of the Haiti earthquake response (McGearty et al., 2012); and in the concept of the UN's transformative agenda as in the IASC real-time evaluations (Cosgrave et al., 2010; Grünewald et al., 2010; Polastro et al., 2011a);

[9] This section is based in part on Beck and Buchanan-Smith (2008)

it can also be found in system-wide evaluations, as in the case of the tsunami (Telford et al., 2006). Joint evaluation can be carried out either for accountability or learning.

6.2.1 Benefits of a joint evaluation

Reasons for doing joint evaluations include the following (see Breier, 2005):

- To see the big picture and to evaluate against that with more of a policy focus
- To avoid the challenge of attribution of impact to any one agency
- For strategic reasons (joint evaluations often have greater credibility and can be useful for advocating change)
- To help participating agencies understand each other's approaches and exchange good practices
- For managerial and financial reasons, for example pooling evaluation resources and capacity
- To reduce transaction costs for developing countries (a joint evaluation may reduce the number of single-agency evaluations all consulting the same stakeholders).

Disadvantages of a joint evaluation include the complexity of the process, which can make it a time-consuming project with high transaction costs and complicated management structures.

Key questions to ask in deciding whether to launch a joint evaluation include the following:

- What is the scope and purpose of the exercise?
- What are the key questions to be addressed?
- Is collaboration the most appropriate approach?

Although the stated purpose of joint EHAs is usually both accountability and learning, in practice participating agencies, for example the Tsunami Evaluation Coalition, have often found them to be most useful for learning.

6.2.2 How to plan and manage a joint evaluation

Being utilisation-focused requires involving the primary intended users of the evaluation in the planning stage that culminates in the ToR. In a joint evaluation, there are many more primary users, and therefore negotiating the ToR is likely to take much longer. This is an essential step, so it is worth planning for the extra time rather than cutting this step short. Indeed, this may be one of the reasons that the ToR for joint evaluations are generally better designed than those for single-agency evaluations (see **Beck and Buchanan-Smith, 2008**). Strong chairing and facilitation skills for this initial planning stage are essential, to manage the negotiations between multiple partners and to ensure that the final ToR are clear and concise rather than a long list of questions that are not prioritised.

Give careful consideration to the management system (see Section 3.7). Two options have been well tested for joint evaluations:

- **Two-tier management system** – a larger steering group that comprises all stakeholder agencies and a smaller management group to manage the process and the evaluation team on a day-to-day basis. The latter can also act as a neutral buffer between the stakeholder agencies and the evaluation team if evaluation findings are particularly contentious (see **Dabelstein, 1996**).

- **Three-tier management system** – a steering group and management group as described above (usually headquarters-based), with an additional in-country reference group or steering committee that builds ownership in the country where the evaluation will take place.

Complex management structures and the high transaction costs of engaging in a joint evaluation can be a barrier to NGO involvement, especially for smaller and national NGOs. Identifying engagement options that are less demanding than those for larger stakeholders may help ensure their involvement.

The pool of evaluators with the technical, political, and interpersonal skills needed to lead joint evaluations is still small. Planning and tendering early will help to secure qualified candidates.

6.2.3 What approach and methods to use

For a large-scale joint evaluation, it may be worth adopting a two-phase approach in which the first phase resembles an extended inception phase. In this phase, the evaluation team carries out the scoping of the evaluation (for example, through preliminary meetings and consultation with the stakeholder agencies), documentation review, and a preliminary visit to the country or region where the fieldwork will take place. This can help focus the evaluation. During this phase, the team would also design the full methodology, which would be considered and approved by the steering group. During the second phase, the full consultation and fieldwork are carried out, the data are analysed, and the findings are written up and presented.

The methods used in a joint evaluation are often similar to those of a single-agency evaluation. With greater resources, however, there are usually greater opportunities to consult the affected population, for example by combining qualitative methods such as focus group discussions and quantitative methods such as field surveys. These need to be carefully planned and sequenced, especially when mixed methods are used.

The OECD/DAC has produced guidance on managing joint evaluations (**OECD/DAC, 2006**), and they are the focus of a chapter in ALNAP's seventh review of humanitarian action (**Beck and Buchanan-Smith, 2008**).

6.2.4 How to ensure follow-up

The large number of stakeholders with different needs in a joint evaluation can make follow-up more challenging than it is for a single-agency evaluation – for example, if recommendations are generic and agencies feel free to pick and choose among them. There are ways around this – for example, ensuring that all recommendations are clearly targeted to groups of users if not to individual agencies. Buy-in at the country level, from the beginning of the evaluation process, is essential for the take-up of findings. Dissemination that is well planned in advance and well resourced is key. It should go well beyond distribution of the report; consider holding workshops to encourage reflection and discussion of the evaluation findings.

Good practice example 19

Achieving buy-in at the country level

A sectoral initiative resulted in a number of inter-agency health evaluations involving UN agencies, NGOs, academics, and recipient governments. Learning from their early joint evaluation experiences, the IHE adapted their approach to achieve buy-in at the country level by setting up an in-country steering committee and a preliminary visit by members of the IHE core working group. Within a year of the evaluation being completed, a planning workshop was held in-country, led by the steering group, to discuss follow-up with all stakeholders. The evaluators returned for this workshop (see **Beck and Buchanan-Smith, 2008**)

As with other evaluations, a management response matrix can help to ensure that evaluation recommendations are followed up. Each agency involved may prepare its own management response, but it is possible to have a joint response.

6.3 Real-time evaluations

In a real-time evaluation (RTE), the primary objective is to provide feedback in a participatory way, during fieldwork, to those executing and managing the humanitarian response (Cosgrave et al., 2009b). There has been a huge growth in the number of RTEs in recent years. However, many of these occurred so long after the intervention began that they could more accurately be called early mid-term evaluations.

6.3.1 Advantages

RTEs offer three potential advantages: timeliness, interactivity, and perspective.

- **Timeliness** – RTEs are carried out when key operational and policy decisions are being taken. They can flag important issues that have been overlooked in the rush to meet immediate needs. For example, the CRS 2010 Pakistan RTE took place only 9 weeks after the start of the agency's response and included a 1-day reflection with staff and partners at different locations during which immediate action plans were drawn up (Hagens and Ishida, 2010).
- **Interactivity** – RTEs involve sustained dialogue with staff, both in the field and at headquarters, and can provide a channel for communication between field staff and headquarters that bypasses bureaucratic roadblocks. For example, the 2010 RTE of IASC's response to flooding in Pakistan included a second visit in which findings, conclusions, and recommendations were discussed with stakeholders in a participatory fashion (Polastro et al., 2011b).
- **Perspective** – An RTE team can approach an emergency from a number of different angles, talking to staff at all levels, in different countries, to the affected population, and to partners and government officials. This can provide a view of the operation that is less influenced by day-to-day problem solving.

6.3.2 Key characteristics

RTEs differ from other humanitarian evaluations in their timing, primary stakeholders, and objective (**Table 21**).

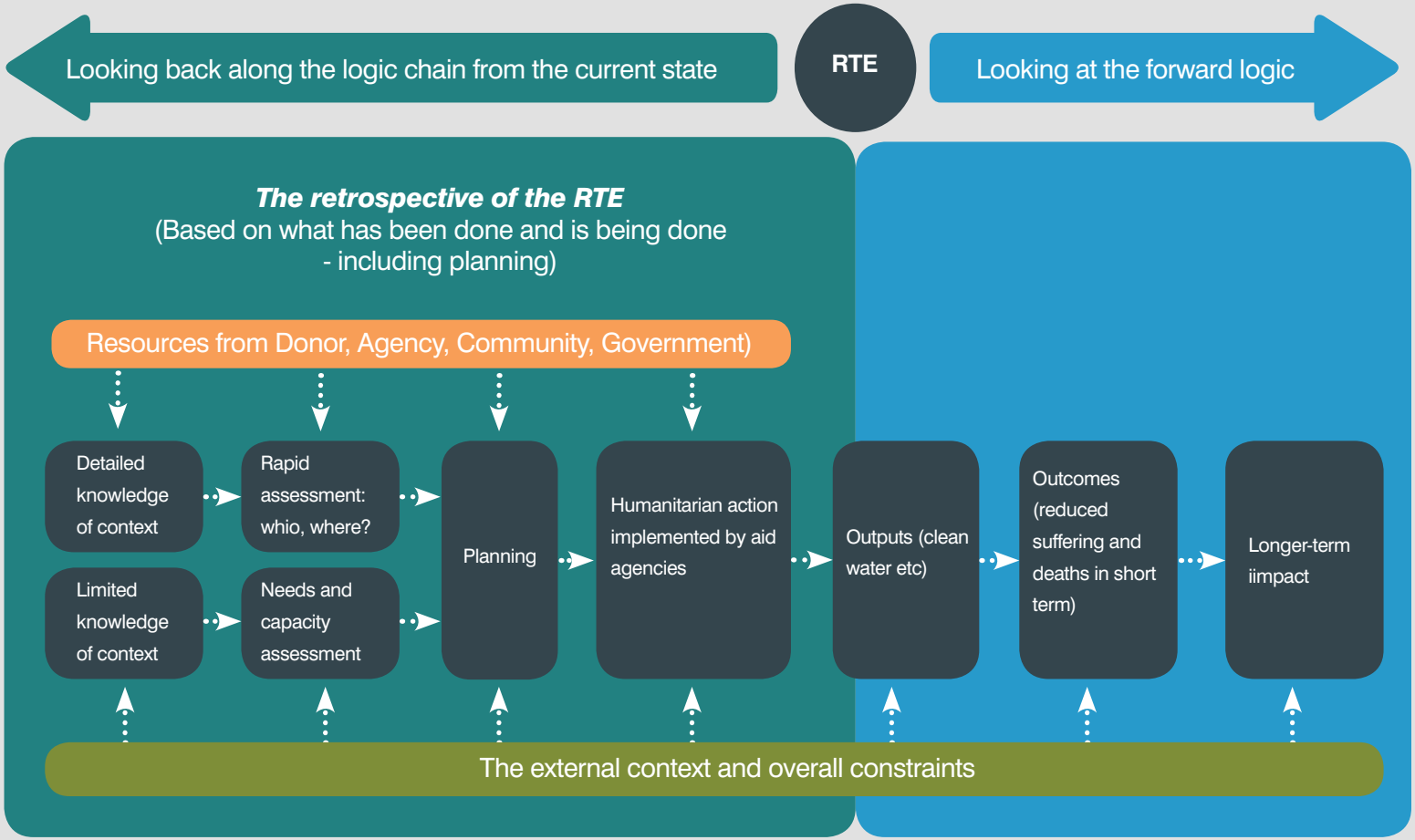
Table 21: Characteristics of a real-time evaluation

Characteristic	Description
Timing	RTEs take place early in an operation; this requires evaluation team members not only to evaluate what has been done but also to look at the likely consequences of what is being done now. RTEs have both forward- and backward-looking components (**Figure 16**). Team members must have sufficient operational experience to understand what the most likely outputs, outcomes, and impacts of current actions are. *Tip: You can get around these time constraints by establishing trigger events for RTEs. Using a generic ToR as a template can speed up the development of a specific ToR, and prequalifying potential evaluators (or using internal evaluators, as UNHCR does) can speed up deployment of the team.* Standard evaluations often allot a month or more after fieldwork to produce a draft report; RTE reports should be finished or nearly finished when the team leaves the field. *Tips: Plan for the team to develop its report before leaving the field. RTE teams may need to remain longer in the field than other evaluation teams; 3 to 4 weeks is probably the minimum. Using an evidence table (Sections 5.2.4 and 5.2.6) can help speed up the team's work.*
Stakeholders	The primary stakeholder for most humanitarian evaluations is an agency's top management. RTEs' primary stakeholders are the field team and those managing the operation from headquarters. The evaluation team must communicate its findings to the team in the field, few of whom would have time to read a traditional evaluation report. *Tips: Include time for debriefings in the field in the evaluation plan. Typically, an RTE requires at least three briefings: one to introduce the evaluation, one mid-fieldwork to raise emerging issues with the field team, and a final briefing.*

Characteristic	Description
Objective	RTEs focus primarily on learning – they are designed to help improve the quality of an on-going response, at a time when there are usually too few outputs to make an accountability-focused evaluation meaningful. *Tip: When you arrive, ask the field team what they would most like to know about their operation.* *Tip: Having a member on the team who is familiar with how the organisation operates will help to improve utilisation.*

Inevitably, an RTE team will consume resources that would otherwise be used to assist the affected population. This creates a tension between what the team can contribute to the operation and the time and resources it requires from the response. Keep the RTE team small in order to simplify logistics and reduce the resource drain. This is especially important if the RTE takes place early in a humanitarian operation.

Figure 16: Forward- and backward-looking aspects of real-time evaluations



Contents
1
2
3
4
5
6
7
Feedback

6.3.3 Methods

Methods used in other EHAs are not always appropriate to RTEs, especially those taking place in the early stage of a response (**Table 22**). Documentary research is less useful than usual; what little documentation exists relates mainly to intentions or inputs, not outcomes. Surveys are usually impossible because of time constraints and lack of a good sampling frame.

RTEs have a strong focus on direct utilisation; a management response matrix is a particularly useful tool for monitoring implementation of the recommendations.

Contents

Table 22: Methods used in real-time evaluations

Method	Use
Focus group discussions	These are more difficult to arrange, as the affected population is usually busy addressing survival needs. A properly managed focus group discussion typically takes two hours or more.
Key informant interviews	These are critical given the operational context, as the interviewees may not have had time to process their experiences in a way that they can readily express without being probed in an interview.
Workshops and after-action reviews	These can be particularly useful, as they provide the field staff with an opportunity to analyse and understand how the operation has developed.
Observations	These are even more important and useful than usual; there will usually be many opportunities to observe implementation in practice.
Beneficiary interviews	These are even more important than usual; they can provide vital and actionable information. The evaluation team will have more opportunities to talk to beneficiaries than to field managers, who need to focus on emerging problems. Interviews take less participant time than focus group discussions and are far simpler to conduct, even in chaotic environments.
Participatory rapid appraisal tools	These can be quite useful, but tools that rely on communal ratings will be constrained by the fact that displaced populations may not yet have a sense of community.

6.4 Impact evaluations

The term impact evaluation has been called a misnomer, as this type of evaluation can more accurately be said to address outputs and outcomes (White, 2009a, p. 8). However, the term remains widely used to refer to evaluation of impact in the results chain. Depending on the programme, this could take place after a few months (for example, for an emergency food aid programme) or after several years (for example, for a livelihood support programme).

Roche (2000, pp. 545–546) provides a useful and widely accepted definition of impact: 'lasting or significant changes – positive or negative, intended or not – in people's lives brought about by a given action or series of actions'. This shifts the focus away from whether the anticipated amounts of aid were delivered or not (effectiveness) to whether people are better off or safer as a result (see Proudlock et al, 2009). However, it should be noted that impact evaluations are not focused on this wider definition, but on the narrower question as to which of the planned outputs and outcomes can be attributed to the intervention.

Impact evaluations are increasingly common in the development sector (Jones et al., 2009, p. v). In the humanitarian sector, conclusive evidence of the impact of a programme is rare, and there are still few examples of evaluations that focus solely on impact. One of the reasons for this appears to be lack of capacity (Proudlock et al., 2009). Interest in impact evaluations is growing, however, under pressure from donors and others to establish the impact of internationally funded humanitarian actions on the lives and livelihoods of the people they serve.

An impact evaluation is likely to be accountability-oriented if it is donor-driven and learning-oriented if it is requested by the implementing organisation. As with other types of evaluations, it is important to clarify the main purpose (Beck, 2011, p. 16). The Feinstein Center has pioneered a participatory approach to impact assessment that creates a learning partnership between the donor and implementing partners (Catley et al., 2009).

More detailed guidance on impact evaluations can be found in (3iE, 2008a, 2008b; White, 2009a, 2009b), Poulos (2006), Roche (1999), Catley et al. (2009), and (Bamberger, 2012; Perrin, 2012; Rogers and Evalution, 2012).

Contents

1 2 3 4 5 6 7

6.4.1 *Joint impact evaluations*

Joint impact evaluations can help ensure that the overall impact of humanitarian action is explored rather than artificially isolating a single agency's work (**Beck, 2009, p. 1**). They also allow comparison of different approaches, which may help to evaluate the relative impact of different interventions. In the humanitarian sector, most agencies planning an impact evaluation are doing so as a joint exercise.

6.4.2 *How to do an impact evaluation*

Impact assessment involves value judgements, for example about which kinds of changes are significant and for whom. It is important to involve the affected population, to gain their perspectives and thus to identify appropriate indicators for impact, which may differ from the indicators selected by other stakeholders (**Proudlock et al., 2009**). Thorough impact evaluations are likely to require longer and more intensive fieldwork than other types of EHA.

A comparative, quantitative approach to impact evaluation is rarely feasible in EHA because the dynamic and fluid environment in which most humanitarian actions take place makes it difficult, if not impossible, to pinpoint the cause of different outcomes (**Proudlock et al., 2009**). More common is a theory-based, qualitative approach that examines a particular case in depth to explain how an intervention could be responsible for particular changes. It triangulates information from a number of different sources, including beneficiaries and key informants. As much evidence as possible should be gathered to support plausible assumptions about causality.

One theory-based method involves three steps:

1. identifying possible causes for the outcomes and impacts of interest
2. identifying the conditions necessary for each possible cause to have an effect
3. establishing whether these conditions are present. This results in a realistic list of potential causes for which the necessary conditions are present (**Scriven, 2008**) as quoted in (**Proudlock et al., 2009, p. 30**).

Contents
1
2
3
4
5
6
7

The lack of baseline data in areas affected by humanitarian crises is a particular challenge to impact evaluation, which may therefore have to depend heavily on the perceptions and judgements of the affected population. They are likely to be the best judge of the changes in their lives, the significance of those changes, and what brought them about. Retrospective baselines can be created based on interviewees' recall of the situation before the crisis or before the intervention.

Impact evaluation could be greatly facilitated by the availability of systematic and reliable monitoring data on outcome indicators, for example malnutrition and morbidity data for health or food distribution programmes. But in practice such data are often lacking, and what data exist are usually quantitative and focused on process and delivery rather than results and impact.

The greatest challenge to impact evaluation is conclusively attributing change to humanitarian action rather than other factors, for example change in the external environment or local people's decisions. Roche concluded that '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' (1999, p. 33).

Additional impact evaluation methods include the following:

- A longitudinal study that captures change in the lives of the affected population over a period of time, for example a year or more, and its causes
- Qualitative methods, such as focus group discussions, group interviews, key informant interviews, and household and individual interviews, either as one-off activities or as part of a longitudinal study
- Quantitative methods, for example formal household surveys or the collection and analysis of quantitative indicators such as market prices (probably accompanied by qualitative methods to establish cause and effect)

Contents
1
2
3
4
5
6
7

A recent consultation on joint humanitarian impact evaluation recommended a mixed-method approach and a greater emphasis on qualitative methods, which would also foster participation. Ideally, mixed-methods support each other in a synergistic way (Adato, 2011). Proudlock et al. conclude 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' (2009, p. 7).

In planning an impact evaluation, consider the following questions:

- How should humanitarian impact be defined – impact on what, and over what timescale?
- 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 context, and how will issues of data, baselines, and timing be addressed?
- How should data on impact be analysed and interpreted, and what role should affected people play in this?

Can this section be improved?

NOTES

NOTES

Please share your feedback:

Help us make the Guide more practical and user-oriented by sharing your feedback on its content and navigation. Feedback can include questions, suggestions, useful resources or practical advice from your own experience. You can share your comments and ideas by clicking on the feedback button at the bottom of any page. This will generate a confidential email to the EHA team in the ALNAP Secretariat. For more information on how you, your team or organisation can engage more actively in piloting the Guide contact us directly at **eha@alnap.org**

OUTPUTS, DISSEMINATION AND TAKE-UP

Contents of this section

7.1 Key outputs 205
7.2 The evaluation report 209
7.3 Circulating and commenting on the draft report 215
7.4 Approving and finalising the evaluation 217
7.5 Dissemination 218
7.6 Facilitating take-up of an evaluation 220
7.7 Evaluation syntheses, thematic reviews, and meta-analyses 221

How to use this section

This section provides guidance on the final phases of the evaluation process. Sections 7.1 and 7.2 address the report and other evaluation outputs. Sections 7.3 and 7.4 cover the review and finalisation stages; Section 7.5 covers dissemination; Section 7.6 discusses ways to encourage take-up; and Section 7.7 reviews other ways of publicising evaluation findings, such as syntheses and meta-analyses.

Tables, figures and boxes

Table 23: Evaluation outputs 206
Box 9: Three audiences for evaluation reports 215

7.1 Key outputs

Evaluation outputs include the inception report, debriefing workshops, advice provided by the evaluation team directly in the field, workshops, the evaluation report, and dissemination events, including workshops. Outputs other than the report (such as briefings at regional headquarters) may need to be considered in the budget and schedule. **Table 23** summarises the main evaluation outputs and their timing and approval processes.

Personal interactions in the field, in briefings, and in workshops are more likely than written reports to lead to learning. Be sure to plan for these, especially in evaluations with a learning focus.

Table 23: Evaluation outputs

Output	Timing	Approval process
Inception report	Not less than 2 weeks prior to the fieldwork to allow time for comments and correction	• Circulation to stakeholders for comments • Meeting between reference or steering group and evaluation team
Advice provided directly in the field	As issues arise in the field and in response to observations	None
Debriefing workshops	Midway through the RTE for evaluation team members to test emerging findings At the end of fieldwork in other types of evaluation	Stakeholders validate or challenge evaluators' findings and conclusions
Debriefing note	At the end of the fieldwork (can be as simple as the presentation given in the debriefing workshop)	Final version should reflect the comments made in the debriefing meeting

Output	Timing	Approval process
Draft evaluation report	Three to four weeks after the end of the fieldwork, longer for more complex reports; may have several iterations	Initial quality check by evaluation manager Circulation to reference group or steering group, key stakeholders, and interviewees for their comments Meeting between reference group or steering group and evaluation team to discuss key issues
Stakeholder workshop	After the first draft report	Stakeholders can develop recommendations and conclusions
Final evaluation report	About 2 weeks after final comments received on penultimate draft	
Video segments on key messages	At the time of presentation of the final report	Review by evaluation manager and reference group or steering group
Management response matrix	After the final report has been submitted	Prepared by the agency ‘s management team
Summary report (key messages from the evaluation)	About 1 week after final report submitted	Review by evaluation manager

Output	Timing	Approval process
Dissemination workshops	At any time after finalisation of the report	None
Brown-bag meetings (short lunch-time presentation of key lessons)	At any time after finalisation of the report	None
Other dissemination activities (such as using findings in trainings)	At any time after submission of final report	None

7.2 The evaluation report

The shorter the report, the more likely it is to be read, but it needs to provide sufficient evidence for the findings, conclusion, and recommendations. The fewer the evaluation questions in the ToR, the easier it will be to balance these two requirements[10].

If the evaluation report is expected to adhere to a particular format, this should be stated in the ToR. This includes details such as fonts, margin sizes, and length.

The ***ALNAP Quality Proforma*** **(ALNAP, 2005)** provides an effective template for an evaluation report, and the United Nations Evaluation Group has provided a quality checklist (**UNEG, 2010**).

Prepare a report outline before fieldwork begins; this will help you structure data collection and analysis to fit the requirements of the final report, thus reducing the amount of work involved in producing the report. It is particularly useful for collating and analysing qualitative data collected in interviews.

[10] The report contents described here are based on the **ALNAP Quality Proforma (ALNAP, 2005)**.

Most reports are structured around the evaluation framework and the evaluation criteria. The most common arrangements involve either chapters for each framework element with sub-chapters for criteria, or chapters for each criterion with sub-chapters for framework elements.

7.2.1 Front matter

Front matter traditionally includes the following:

- **Title page** – include a date.
- **Data page (optional)** – this can include the administrative details for the evaluation, acknowledgements, and a suggested citation for the evaluation.
- **Executive summary** – this may include a summary of the main recommendations. It should be omitted in early drafts of the report to avoid getting comments on the summary rather than the full report.
- **Table of contents** – if appropriate, also include lists of tables and figures. Most word processing software can automatically generate these if headings and captions are formatted appropriately (for example, with Microsoft Word styles).

- **List of acronyms** – this is appropriate if the document contains more than a few acronyms. If it also contains unfamiliar terms (such as local words for a particular geographical feature or water source), provide a separate list of these terms with definitions.
- **Map(s)** – either generate your own maps using mapping software or look for a non-copyright map.

Use acronyms only when a term occurs so frequently that having a short form is helpful to readers, or when it is more recognisable in its acronym form. Always spell out acronyms and define unfamiliar terms on first use in the text, even if you think they will be familiar to many readers. These practices not only make a document more readable but also help emphasise that the report was written for all stakeholders, not just agency or sector insiders.

7.2.2 The main text

The main text of a report usually has the following elements:

- **An introduction** that sets out the scope and purpose of the evaluation, the team composition, and the structure of the report.
- **A methodology section** that describes the main methods used, the constraints encountered, and any biases in the evaluation team. This may be part of the introduction and may be a simple summary that refers readers to a more detailed discussion in an annex.
- **A context chapter.** This may be part of the introduction, and may be supported by a chronology in an annexe, but it should be thorough. Context is a key issue in humanitarian action: whether actions are appropriate, effective, efficient, or coherent depends very much on the context.
- **The principal chapters**, presenting the evidence and findings, organised by the evaluation criteria or some other framework. Sometimes conclusions and recommendations are presented at the end of each chapter, and sometimes they are presented in one or two separate chapters at the end of the report.

Presenting the conclusions and recommendations at the end of each chapter emphasises (and helps ensure) their grounding in the relevant evidence, findings, and conclusions.

Good practice example 20

Using a table to summarise multiple actions by multiple partners

An evaluation of the Pakistan floods recovery work for the CBHA summarises in a one-page table the varying experiences of six agencies with cash, vouchers, and in-kind support across three livelihood sectors (**Leturque et al., 2012, p. 45**). The table also summarises the theoretical advantages and disadvantages of the three modalities.

	Theorectical pros and cons	Result 1 – Agric inputs	Result 2 – Livestock	Result 3 – Business Assets
Conditional Cash Grants	Pros: • Agency can influence recipient expenditure • Recipient may have other priorities Cons: • Administratively heavy (expenditure monitoring before payment)	**CARE International** Successes: Flexibility of conditional cash grants limits seasonality-related risks. Beneficiaries able to adapt their input purchase to match the timing of cash transfers.	**CARE International** Although beneficiaries have the option, Cash grant almost always used for agriculture inputs rather than for livestock.	**CARE International, Concern, IRC, Oxfam** Successes and challenges: Indications that about 1/2 of more of the grant was spent on other purposes.

	Theorectical pros and cons	Result 1 – Agric inputs	Result 2 – Livestock	Result 3 – Business Assets
Vouchers	Pros: • Better control over achievement of market objectives Cons: • Important planning required • Stimulate local business • Excludes some traders • Selected trader may inflate prices • Limit recipient choices	**ACF, Oxfam** Successes: • Targeted inputs received. • Timely execution. Challenges: • Beneficiaries reported distorted prices by selected suppliers.	Not used by CBHA partners	**SC** Challenges: • Feedback from beneficiaries that they would have preferred cash grants. • Limited choices and price inflated by shortlisted providers.
In-kind	Pros: • More targeted objectives • Control over quality of inputs Cons: • Risk associated with transport • High logistic costs • Limit beneficiary choices	**Concern, IRC, Oxfam, SC** Successes: • Significant savings on bulk purchase. • No major delivery delays reported.	**Concern, IRC, Oxfam, SC** Successes: • Savings on bulk purchases Challenges: • Late delivery (Oxfam) • High mortality (late delivery, transport) • Animal feed packages useful but not sustainable • Set ration difficult to control • Training required and timeline of the project maladapted to capacity building.	Vocational Training: not analysed in this evaluation

Tables are only one way to communicate information quickly. Carefully chosen photographs, well-constructed charts, and other figures can also communicate key findings. Tufte (1990; 2001; 1997) provides good advice on the presentation of visual information.

7.2.3 Recommendations

Recommendations are only useful if they are:

- **Specific** – it must be clear exactly what is being recommended.
- **Related to verifiable actions** – it should be possible to tell whether the recommendation has been implemented or not.
- **Directed** – the person or entity responsible for implementing the recommendation should be identified; responsibility may be further clarified in a management response to the report.
- **Practicable** – recommendations can involve new or unusual ways of doing things, but they should bear resources and other constraints in mind.
- **Time-bound** – a timetable for implementing the recommendations should be given wherever possible.
- **Consistent** – recommendations should not contradict or seem to contradict each other.
- **Prioritised** – it should be clear which recommendations are of primary concern and which are secondary.
- **Economical** – the recommended actions should clearly deliver benefits in proportion to their costs.

Recommendations should also be limited in number. The fewer recommendations a report contains, the easier it is for the client to use it. However, an experienced evaluator may notice dozens of performance issues and has an ethical duty to raise them. Various strategies can help resolve this conflict:

- Offer a general recommendation, and then provide details of how this could be implemented by different actors.
- Make minor recommendations orally and put them in an annexe.
- Rank recommendations by importance.
- Group recommendations by the target of the recommendation.

7.2.4 Annexes

Annexes are used for supporting elements and for detail that would clutter the main report. They help readers get a sense of the quality of the evaluation by demonstrating the extent of the fieldwork and the range of methods used. The annexes may include the following:

- **ToR** – by convention this is the first annexe to an evaluation report.
- **Chronology of the issue or action being evaluated** – this is essential in any evaluations in which timeliness is a criterion.
- **List of people met or consulted** – this demonstrates how extensively the team consulted different actors, and helps readers form an opinion of the likely reliability of the findings.
- **Team itinerary** – this also provides readers with a view of the extent of the team's research.
- ***Team biographies*** – these give readers a sense of team members' experience, as well as any biases they may bring to the evaluation.
- ***Discussion of methodology*** – this sets out the methods used in detail and acknowledges any constraints and the limitations they may have. This is normally supplemented by examples of the different interview instruments and survey forms used.
- **Management response matrix** – this may be attached to the report as an annexe; more frequently, it is printed as a separate document.
- **Other annexes** – these could address topics for which a detailed discussion would be out of place in the main report, or present results from specific methods (such as a summary of the responses to an online survey).
- **Bibliography** – this may include key sources consulted in addition to the sources cited in the evaluation report.

A current trend is to make most or all of the annexes available on a website rather than publishing them with the report itself. Unfortunately, it is often not possible to find these annexes after a few years. For example, the annexes to the 2011 IASC Somalia evaluation (**Polastro et al., 2011a**) can no longer be found on the UN website for Somalia.

Even if the annexes will not be part of most hard copies of the report, produce an electronic copy complete with annexes that can be posted on the client's website.

Box 9: Three audiences for evaluation reports

Every evaluation report has three broad audiences.

1. Managers don't have the time to ready a lengthy report. Their needs are often best served by a good executive summary a few pages long. (Many executive summaries are far too long to serve their purpose.)
2. Stakeholders can read the full report.
3. Technical specialists rely on the detailed information provided in the annexes.

When you write each element of the report, think of the audience for which that element will be of the greatest interest.

7.3 Circulating and commenting on the draft report

Draft evaluation reports are usually circulated by the evaluation manager to key stakeholders and possibly to all interviewees. Giving interviewees an opportunity to respond to any errors of fact or understanding is an effective quality control measure. It requires the evaluation team to record email addresses for all interviewees.

Circulating the report also enables key stakeholders to call attention to any errors of fact, understanding, or analysis in the report. This can help to promote stakeholder ownership of the evaluation and encourage utilisation.

Allow sufficient time – usually 1 to 3 weeks – for this review. Stakeholders who receive the draft may need to circulate it within their agencies and then collate the comments they receive.

Circulating reports as word processing (usually Microsoft Word) documents creates problems. The temptation is to embed comments as tracked changes, which become difficult to follow if more than two or three people leave comments in the same document. Different user settings can also cause changes in pagination and layout, which can create confusion. It is better to circulate the report as a PDF file, which provides a fixed layout and offers more legible commenting options.

Number paragraphs in the draft report to make it easier to write comments referring to specific text passages.

Collate comments and work through them systematically, so that all comments on a particular section are considered at the same time.

Good practice example 21

Collating comments in a transparent way

An evaluation of the Pakistan floods recovery work for the CBHA summarises in a one-page table the varying experiences of six agencies with cash, vouchers, and in-kind support across three livelihood sectors (**Leturque et al., 2012, p. 45**). The table also summarises the theoretical advantages and disadvantages of the three modalities.

Page 42 para: 151 – *One very successful innovation in this response was the appointment of a very high profile UN special envoy, former president Bill Clinton. While initially this choice may have been motivated by a desire to have an Envoy who could shake loose the donor's purse strings (Clinton, 2006) it proved very successful when the focus changed to promoting the effectiveness of the response.*

UN Agency 1 – This point should also mention the creation and role of the Global Consortium on Tsunami Recovery chaired by Bill Clinton, and the innovative coordinating and communication role with a line to the IASC–UNDG.

UN Agency 2 – The coordination report also lauds the appointment of a Special Representative for the Tsunami – the OCHA DERC (Margareta Wahlstrom). It was considered another positive coordination function. Don't know why you chose not to pick that up…but of course that's the team's prerogative.

Reviewer A – It's not clear on what basis the conclusion that the Special Envoy was helpful is being made.

NGO 1 – It might be useful to spell out more specifically what made Clinton's appointment a "success".

Letting reviewers know that you will collate and circulate their comments encourages more considered comments.

It is good practice for the evaluation team to indicate how they have responded to reviewers' comments by adding brief notes such as the following:

- Corrected
- Nuanced
- Detail added
- Deleted
- Not accepted [with a detailed reason]

7.4 Approving and finalising the evaluation

The evaluation report is the main output from an evaluation. It should do the following:

- Answer the evaluation questions.
- Provide evidence to support its conclusions and recommendations.
- Present clear recommendations and identify who should implement them.
- Be coherent and free from internal contradictions.
- Use the methods specified in the ToR and inception report, and if not, clearly explain why.
- Meet any layout and formatting requirements.
- Contain sufficient information about process and methods to enable readers to judge its reliability.

Readers of the evaluation report will make their own assessment of the quality of the evaluation and of the authority of the recommendations. A good report reflects well on the evaluation manager and the evaluation team; a bad report reflects poorly.

7.5 Dissemination

All too often, planning for dissemination of the evaluation findings and recommendations is left until after the evaluation report has been finalised. But if you are serious about utilisation, the entire evaluation – including dissemination – must be planned and budgeted with this in mind (Section 2). Delivery of the final evaluation report is not the end of the evaluation, it is simply the beginning of the dissemination phase.

Dissemination should be guided by a communication strategy indicating the following:

- The key groups to whom the evaluation findings and recommendations need to be communicated
- The best means for communicating with these groups
- Who is responsible for the communication

- Most people learn and remember more from hearing and discussing than from reading. Merely circulating the report is not enough. A wealth of additional options exist for communicating evaluation results:
  - Targeted meetings can be held between the evaluation team and intended users. If these are designed workshop-style, and customised to suit the needs of each user group, they may encourage greater engagement by users.
  - One-to-one briefings can be offered to key users by the evaluation team or the evaluation manager.
  - Short, accessible summaries or briefing documents are more likely to be read by busy decision-makers than a long evaluation report.
  - Short emails can communicate findings, conclusions, or recommendations to key users.
  - The evaluation or its summary can be translated into local languages and disseminated to local stakeholders.

- Members of the evaluation team can return to the site of the evaluation to report to key stakeholders, such as programme staff and members of the affected population.
- Findings and recommendations can be communicated using video footage from the evaluation or recorded interviews with the evaluation team.
- Podcasts (audio recordings that can be listened to on MP3 players and computers) discussing key points from the evaluation can be made available.

Good practice example 22

Using video to disseminate key messages

A number of organisations are experimenting with the use of videos to accompany their evaluations – for example, UNHCR, the Red Cross, and Groupe URD. These can provide powerful feedback on key issues in the words of beneficiaries, and are more likely to be remembered than are long evaluation reports. The Groupe URD video on the Haiti real-time evaluation can be viewed online at http://vimeo.com/15198053.

Social media show promise as ways to communicate evaluation results (**Kaplan and Haenlien, 2010**) – including collaborative projects like wikis, social networking sites like LinkedIn, blogs, and content communities like YouTube. For example:

- Details of the progress of the UNHCR Age and Gender Diversity Mainstreaming evaluation for Colombia (Mendoza and Thomas, 2009) were updated on a blog called **It Begins with Me. It Begins with You. It Begins with Us: Thoughts and Actions Around Age, Gender and Diversity Mainstreaming**.
- The European Evaluation Society conducts discussions on LinkedIn.

Evaluation updates can be provided regularly via microblogging. Probably the best known such service is Twitter, which limits messages to 140 characters – a format better suited to brief updates on a dynamic process than to detailed discussions. Social media continue to evolve rapidly, and new forms and uses are continually emerging.

7.6 Facilitating take-up of an evaluation

Involving the primary stakeholders throughout the evaluation process is key to utilisation. A well-planned and well-funded dissemination phase is also important. Other ways of facilitating take-up include the following:

- Involve some of the key users in formulating the recommendations.
- Clearly allocate responsibility for follow-up. A formal management response matrix is one way of doing this.
- Build the findings of the evaluation into training materials, for example as case studies.

Good practice example 23

Involving key stakeholders throughout the evaluation

The ECB evaluation of the response to the Yogyakarta earthquake (Wilson et al., 2007) had a strong utilisation focus. Key aspects included the following:

- discussion at the outset with in-country steering committee members regarding what they wanted from the evaluation
- dialogue between the team leader and the steering committee before the team leader arrived in-country
- discussion between the evaluation team and the steering committee about evaluation methods – for example, the most appropriate interpretation of the DAC criteria

The evaluation team presented summary evaluation findings to the steering committee and to field staff so that these primary stakeholders could work with the team to draw up conclusions and recommendations. A final meeting was held between the evaluation team, members of the steering committee, government officials, and local people to review and amend the preliminary conclusions and to make further recommendations.

7.7 Evaluation syntheses, thematic reviews, and meta-analyses

The length and detail of an evaluation report can discourage decision-makers from reading it and incorporating its lessons into future interventions. A popular and effective way of ensuring that key findings reach their target audience is to produce a synthesis of past evaluations, for example as a short and accessible lessons-learned paper. A number of organisations have carried out such meta-reviews, sometimes on thematic grounds, and these are usually welcomed by potential users. See good practice example 24.

Good practice example 24

Synthesising lessons learned

In 2008, in response to the earthquake in China, ALNAP produced a paper on humanitarian activities after earthquakes that distilled the lessons from a number of evaluations of earthquake responses (**Cosgrave, 2008**). In the days following the 2010 earthquake in Haiti, the paper was downloaded more than 3,500 times.

CARE carried out a meta-review of evaluations and after-action reviews from 24 emergency responses, exploring the extent to which CARE had internalised recommendations and lessons from the evaluations. The aim was to reflect on how CARE might make more effective use of its evaluation findings (**Oliver, 2007**).

World Vision has produced a compact disc with more than 50 lessons-learned summaries across a range of sectors (**Montbiot, 2006**).

Can this section be improved?

NOTES

NOTES

Please share your feedback:

Help us make the Guide more practical and user-oriented by sharing your feedback on its content and navigation. Feedback can include questions, suggestions, useful resources or practical advice from your own experience. You can share your comments and ideas by clicking on the feedback button at the bottom of any page. This will generate a confidential email to the EHA team in the ALNAP Secretariat. For more information on how you, your team or organisation can engage more actively in piloting the Guide contact us directly at **eha@alnap.org**

ALNAP

c/o Overseas Development Institute
111 Westminster Bridge Road
London SE1 7JD
United Kingdom

Tel +44 (0)20 7922 0300
Fax +44 (0)20 7922 0399
Email **alnap@alnap.org**