

Evaluation of Protection in Humanitarian Action



ALNAP is a global network of NGOs, UN agencies, members of the Red Cross/Crescent Movement, donors, academics and consultants dedicated to learning how to improve response to humanitarian crises.

www.alnap.org

Authors

Ian Christoplos and Neil Dillon with Francesca Bonino*

Acknowledgements

Experts consulted and interviewed in developing this Guide: Jon Bennett, Margie Buchanan-Smith, Vanessa Corlazzoli, John Cosgrave, Jamie Khalil, Mark Rogers and Seb Taylor.

External reviewers: Luis Enrique Eguren Fernandez, Guilhem Ravier, Kathrine Starup and Yanya Viskovich

ALNAP internal review: Paul Knox Clarke, John Mitchell, Luz Saavedra and Alexandra Warner

Research support: Sian Cook

The views contained in this report do not necessarily reflect those of the ALNAP Members.



ALNAP (2018) Evaluation of Protection in Humanitarian Action. ALNAP Guide. London: ALNAP/ODI.

© ALNAP/ODI 2018. This work is licensed under a Creative Commons Attribution-Non Commercial Licence (CC BY-NC 4.0).

ISBN 978-1-910454-70-1

Communications management by Maria Gili, Tim Harcourt-Powell and Cara Casey-Boyce

Design by Human After All

* At the time of writing, Francesca Bonino was a Research Fellow at ALNAP. She now works as a Senior Evaluation Officer with UNHCR.

Contents

Introduction

5	Acronyms
7	1 Why a Guide on evaluating protection in humanitarian action?
11	2 Main features of this Guide
14	3 Brief orientation on protection in humanitarian action

A / Initiating the evaluation of protection

24	4 Protection-specific evaluability conditions and opportunities to promote utility
27	5 Evaluation questions and the intervention logic
40	6 Indicators
46	7 Selecting approaches, designs and methods

B / Data management: legal and ethical implications

58	8 Ensuring that the evaluation is carried out in a protective and conflict-sensitive manner
67	9 Data sources and constraints on data management
76	10 Data on the less tangible and harder-to-measure dimensions of protection in humanitarian action

C / Analysis

- 85 **11** The starting point for analysis: Revisiting the intervention logic
- 88 **12** Analysing causality, attribution and contribution
- 93 **13** Insights from evaluating advocacy and other initiatives intended to influence the protection environment

Toolkits

- 102 **Toolkit item #1:** Evaluability checklists for the evaluation of protection
- 110 **Toolkit item #2:** Evaluative rubrics
- 114 **Toolkit item #3:** A partial menu of evaluation approaches and designs
- 125 **Toolkit item #4:** ODI/RAPID approach to theories of change
- 127 **Toolkit item #5:** ODI/RAPID influence and interest matrix
- 129 **Toolkit item #6:** Overview of contribution analysis

The bibliography is available at alnap.org/help-library/eval-protection-biblio.

Acronyms

ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
CA	Contribution Analysis
CHS	Core Humanitarian Standard
DE	Development Evaluation
DFID	Department for International Development
EHA	Evaluation of Humanitarian Action
FGD	Focus group discussion
GBC	Global Protection Cluster
GBV	Gender-based violence
ICRC	International Committee of the Red Cross
IDP	Internally displaced person(s)
MSC	Most Significant Change
M&E	Monitoring and Evaluation
OM	Outcome Mapping
PIA	Participatory Impact Assessment
QCA	Qualitative Comparative Analysis
RDD	Regression Discontinuity Designs
RE	Realist Evaluation
RTE	Real-Time Evaluation
SCM	Success Case Method
SNA	Social Network Analysis
TBE	Theory-Based Approaches to Evaluation
ToC	Theory of Change
U-FE	Utilisation-Focused Evaluation

Notes

1 Why a Guide on evaluating protection in humanitarian action?

People affected by crisis face a wide range of risks. All too often, armed conflict leads to the deliberate targeting of civilians, the violation of human rights, or discrimination in access to services. Natural disasters and displacement bring their own risks, including the increased threat of sexual and gender-based violence (GBV) faced by internally displaced persons (IDPs), or the violations of the rights of the child as enshrined in international law.

The international humanitarian system has increasingly recognised the gravity and centrality of these risks, as witnessed, among others, by Slim and Bonwick (2005) and Niland et al. (2005). The number of actors seeking to reduce protection risks for populations affected by crisis has steadily risen, and the system has increasingly developed sector-wide policies, standards and tools for protection. Key advances include:

- The 2009 ICRC Professional Standards for Protection Work, and their subsequent revisions in 2013 and 2018, which provide a set of commonly agreed standards applicable to all humanitarian and human rights actors doing protection work in conflict and other situations of violence.
- The 2012 Minimum Standards for Protection Mainstreaming, which provide practical assistance to humanitarian actors to mainstream protection in the assessment, design, implementation, monitoring and evaluation of humanitarian programmes, projects and activities.
- The 2013 Inter-Agency Standing Committee (IASC) Statement on the Centrality of Protection, through which the IASC Principals committed to supporting IASC members and field-level teams to ensure protection is placed at the centre of humanitarian action.
- The 2016 IASC Policy on Protection in Humanitarian Action, which notes that protection is a shared responsibility at the forefront of humanitarian action and requiring system-wide approaches.

- The 2017 Protection Mainstreaming Toolkit, developed by the Global Protection Cluster (GPC), to provide guidance to both protection and non-protection specialists on mainstreaming protection throughout the humanitarian programme cycle.
- The 2017 ICRC/Brussels Privacy Hub Handbook on Data Protection in Humanitarian Action.
- The revised Sphere Handbook for Humanitarian Response, which includes an updated chapter on protection in humanitarian aid.
- The Core Humanitarian Standard (CHS) on Quality and Accountability, whose nine principles for improving quality and accountability of all humanitarian assistance cover the four elements of protection mainstreaming as identified by GPC (2014).

In contrast to the attention paid to the planning, coordination and conduct of protection work, relatively little has been published about evaluating protection. As noted by Bonino (2014), evaluation guidance focusing on protection in humanitarian action is limited, fragmented and confined to specific programming manuals that often give limited guidance on the overall challenge of looking at protection. It is therefore unsurprising that evaluations of humanitarian action take account of protection in a dispersed and inconsistent manner (Bonino, 2014; Reichhold and Binder, 2013).

At its core, this reflects a fundamental definitional problem for protection work. Protection is variously understood as an activity, as an approach or lens through which to understand humanitarian action, or as a goal or objective of such action (Murray and Landry, 2013).

Moreover, beyond the question of defining protection itself, evaluators face at least three further challenges that, although familiar from other areas of evaluative activity, are greatly exacerbated by the nature of protection work.

Variation in protection approaches: Protection programme models typically vary between protection-specific activities, integration of protection concerns in sector-specific projects, and efforts to mainstream protection across all humanitarian activity.¹ This makes it difficult to understand what success looks like. How should evaluators understand and interrogate the theory of change (ToC) behind a protection intervention? Should they draw the line

between protection objectives, such as reducing GBV in the household, and broader objectives, such as strengthening resilience at the household level? If so, how?

Understanding cause–effect relationships: This common evaluation challenge is further complicated by the complexity of protection work. Reducing a population’s exposure to risk always involves actions on the part of a wide range of stakeholders, including states, non-state duty bearers and humanitarian actors. The influence of external actors on the results of a given intervention is therefore greatly amplified in protection activities, making causal attribution harder than in other cases. Moreover, the nature of protection outcomes often involves behaviour change within and across a community. Causal relationships between changes in community-wide behaviours and intervention activities are notoriously difficult to isolate with certainty – even if it is assumed such changes can be reliably measured over time.

Collecting data: Good data collection always faces logistical challenges in crisis contexts and there are additional concerns in relation to evaluations looking at protection outcomes. First, it often becomes essential to access and manage very sensitive data, sometimes drawn from communities in conflict. It is therefore of paramount importance to have an ethically sound approach to data collection, storage and analysis. In addition, protection outcomes sometimes include reductions in perceived risk among affected populations, or improvements to protection environments advocated for on their behalf. Outcomes such as these require evaluators to collect robust data on often intangible outcomes, such as perceptions or attitudinal shifts at organisational, governmental and/or societal level. Although the evaluation community is familiar with these challenges, these are compounded by their co-existence and overlap in already challenging evaluative environments.

This Guide, a companion to ALNAP (2016) Evaluation of Humanitarian Action Guide (hereafter the ALNAP EHA Guide), seeks to address these challenges. It offers protection-specific insights for evaluators and evaluation commissioners across the ALNAP Membership. It does not attempt to define protection but is rather intended as support for evaluators and evaluation managers involved in analysing interventions that take their points of departure from a variety of definitions.

The pilot process for this Guide

After a period of piloting and feedback by ALNAP Network Members in 2016, this companion Guide was updated in 2017. Feedback was gathered through three workshops;² a review of 25 protection-related evaluations and 32 policy documents (listed in the Bibliography); and the piloting of the Guide during an end-of-project evaluation commissioned by the Danish Refugee Council (DRC, 2017) in Southern Turkey. The Guide was also cross-referenced against the ALNAP EHA Guide³ and the revised ICRC Protection Standards (ICRC, 2018).

Objectives

The scoping phase for this Guide (Bonino, 2014) found that existing theory and practice do not adequately address the evaluation challenges that protection work raises.

This Guide provides evaluation teams and commissioning staff with tools to navigate these challenges. More specifically, the objectives are to:

- Focus on the decision-making processes related to evaluating protection in humanitarian action and presenting the critical decision points in an evaluation whose focus includes protection.
- Describe some of the trade-offs required and options available to evaluators and evaluation commissioning offices in preparing for an evaluation, selecting approaches and methods and gathering data.
- Offer practical insights, tools and approaches that can be used in evaluating protection in humanitarian action.

2 Main features of this Guide

Target users: The Guide addresses a broad audience of evaluators, staff who commission evaluations and staff in evaluation management and advisory roles (in agencies with and without a specific protection mandate).

Agency staff in protection programming advisory and support roles who are not ‘evaluation experts’ are often called upon to support the management of evaluations, comment on evaluation terms of reference (ToR), review evaluation proposals, be part of mixed evaluation teams and provide technical support to evaluations that are commissioned specifically to look at protection programming in a given humanitarian context. This Guide is therefore designed to be accessible to staff without an evaluation background. It is recommended, however, that this target group also consult the more comprehensive ALNAP EHA Guide.

Some familiarity with key terms and concepts in EHA and in protection is assumed. These terms are presented in [Section 3](#).

The Guide can be read from start to finish, but we suggest users **focus on different modules** depending which stage of the evaluation process is most relevant to them:

Module A	Module B	Module C
Initiating the evaluation of protection in humanitarian action	Data management: legal and ethical implications	Analysis

This Guide is designed to:

- Focus specifically on decision-making processes and options in an evaluation process.

- Address evaluators, evaluation managers and staff involved in commissioning evaluations working in both ‘protection specialist’ and ‘generalist’ positions within agencies, with or without a specific protection mandate in different operational environments, as well as to independent evaluators.

The Guide is not designed to:

- Provide comprehensive guidance covering all the steps in a programming cycle. It focuses on specific decisions related to initiating, scoping, designing and undertaking evaluations that look at protection.
- Duplicate the ALNAP EHA Guide, which remains ALNAP’s main entry point to and reference text on humanitarian evaluation.

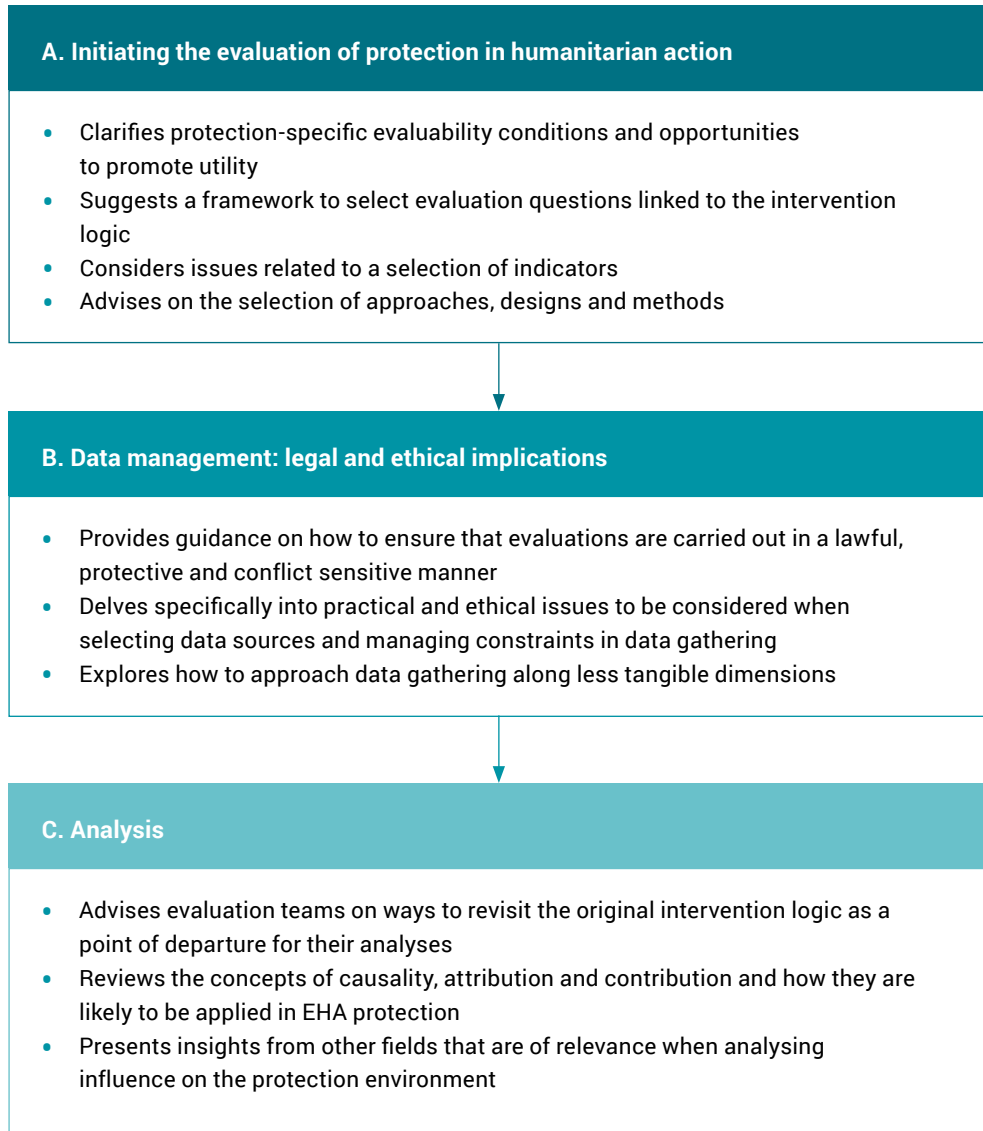
Features of the Evaluation of Protection in Humanitarian Action Guide

To improve accessibility, this Guide features:

- A **detailed content map** presented as a flowchart that features the Guide’s components and indicates to which stage of a generic evaluation (and pre-evaluation) process they refer.
- A number of boxes featuring **evaluator insights**, or nuggets from evaluation practitioners reflecting on how they used a specific tool or framework presented in the Guide.
- A **toolkit section** to describe in more detail selected tools and approaches mentioned in the main text. Further tools are available at www.globalprotectioncluster.org/en/tools-and-guidance/essential-protection-guidance-and-tools.html

Whenever a specific tool or another part of the Guide is referred to, it is underlined in teal.

Figure 1: Evaluating humanitarian protection guide flowchart



3 Brief orientation on protection in humanitarian action

The IASC 2013 Statement on the Centrality of Protection affirms that ‘all humanitarian actors have a responsibility to place protection at the centre of humanitarian action’ (IASC, 2013: 1). This was re-affirmed in the IASC’s 2016 Policy on Protection in Humanitarian Action. As such, all evaluations have a reason to consider protection when assessing agency actions.

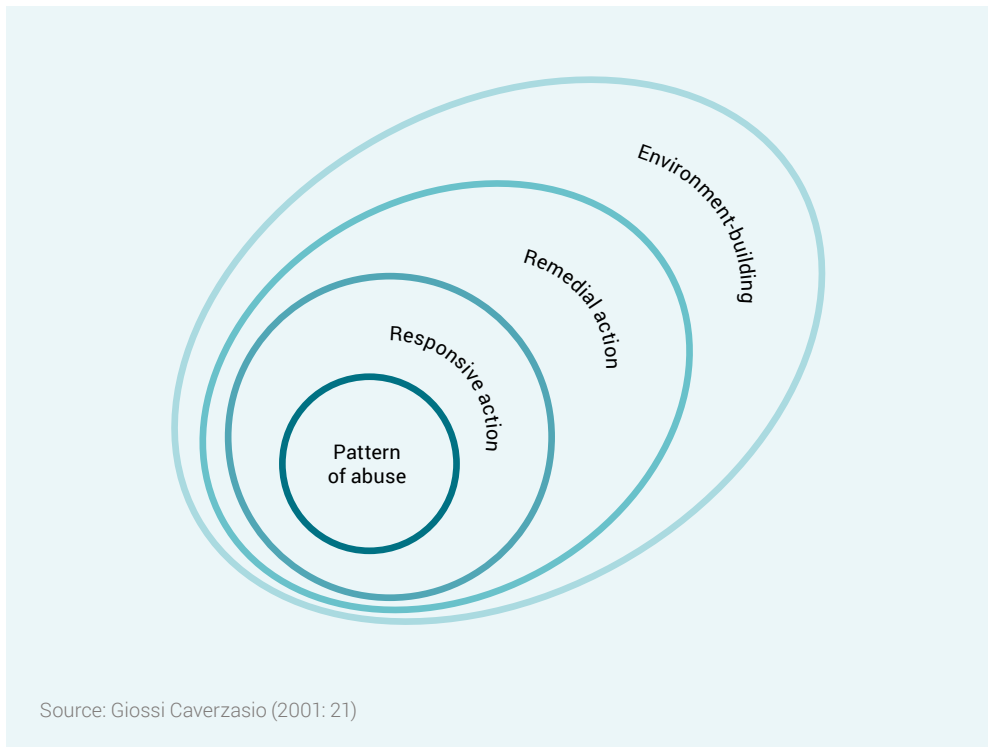
Nevertheless, agency mandates and priorities regarding protection have always differed significantly. This variation has increased in recent years, with an ‘increase in the number and diversity of humanitarian and human rights actors involved in promoting the protection of those at risk of violations or abuses in armed conflict and other violence’ (ICRC, 2018). The definition of the **concept of ‘protection’ is itself multifaceted** and defies clear categorisation and measurement. This has implications for the overall scope of evaluations and the selection of analytical frameworks.

Acknowledging these issues, the International Committee of the Red Cross (ICRC) proposed using the so-called **egg framework on protection** (Figure 2) as a way to show the relations among the different strands of protection work in humanitarian contexts. This framework specifies three interdependent but non-hierarchical families of protection actions (Giossi Caverzasio, 2001: 21-24). Evaluators may be asked to look at activities, services and expected results in all three of them:

- **Responsive actions** to stop, prevent the recurrence, or alleviate the immediate effects of an emerging or established pattern of abuse.
- **Remedial actions** being undertaken after abuse has occurred to restore people’s dignity and ensure adequate living conditions.
- **Environment-building actions** to foster a political, social, cultural, institutional and legislative environment that enables or encourages national authorities to fulfil their obligations and respect individual rights.

As more attention is paid to strengthening national and local institutional capacities to undertake protection actions (beyond humanitarian action), **environment-building** action has come to encompass an expanding range of support that recognises and seeks to reinforce communities' generally under-acknowledged role and address the central role of the state (positive or negative) in protection.

Figure 2: The egg framework on protection



The contexts in which protection actions are carried out have major implications for the selection of evaluation indicators for any given approach. The selection of evaluation indicators largely depends on the type of protection programming being evaluated, and the ways in which a particular programme or agency works with protection. Major factors include:

- the willingness and capacity of the state and the authorities to respond to protection risks and violations
- the capacity of civilian communities to help themselves and their space to act
- the agency's capacity to respond
- the risk the action would create for the security of the civilian population⁴
- the political risks of the action for the agency's security and access
- the duration of the action
- the agency's experience with similar actions in a given setting
- the activities and mandates of other actors.

A study commissioned by the Global Protection Cluster (GPC) neatly captures the multifaceted nature of protection in the context of humanitarian action:

Protection defies neat labelling because it is at the same time the **goal** underlying the whole humanitarian response (the reason for humanitarian action), an **approach** or **lens** on the humanitarian response (a way of understanding all dimensions of humanitarian endeavour), and a more narrowly-defined **family of activities** that aim to prevent and mitigate threats to vulnerable persons. (Murray and Landry, 2013: 4; emphasis in the original)

This complicates the evaluator's tasks of identifying indicators and tailoring methods to assess and judge intended results. It also creates a range of challenges in delimiting and describing what the evaluation needs to look at.

First, when asked to 'evaluate protection', a specific set of protection results, or a protection component of a larger programme or intervention, it is essential to clarify the types of protection included in the intervention, including how the agency is using the concept of protection. Where different

areas of protection and perceived priorities are combined in a given intervention, it is important for evaluators to revisit how these have been delineated.

Second, depending on the purpose of the evaluation, the questions it asks and the orientation of those who commissioned it, evaluators may be asked to look at the scope of protection in humanitarian action such as:

- an overarching theme of analysis for a whole response in a given humanitarian crisis or conflict context
- a specific issue that cuts across different (sectoral) programming areas and interventions
- a primary line of enquiry in an evaluation looking at sector-specific results in a dedicated area of programming (e.g. child protection, GBV and protection against sexual exploitation and abuse)

Third, there can be evaluation scenarios where the ToR do not mention 'protection' as such, but where protection is nonetheless an implicit focus due to its centrality, as outlined in the 2013 IASC Statement on the Centrality of Protection, and elsewhere. Indeed, IASC (2016) 'underlines the needs to implement [the commitment to protection] in all aspects of humanitarian action and across the Humanitarian Programme Cycle (HPC) (IASC, 2016: 2). In such cases, the evaluation team may need to tease out the protective features in a programme that can be inferred from, for example, 'do no harm' measures or the safety and accessibility of the service or assistance provided.

Key to design features



Good practice example



Keep in mind



Evaluator's insight



Tables in this colour represent content.



Tables in this colour present navigation aids for the reader.

Endnotes

- 1 See ECHO (2016: 13-16) for a clear explanation of the differences between these three approaches; and p.68 for a decision tree to choose between programme models on the basis of a protection-risk analysis.
- 2 In December 2016, ALNAP and InterAction co-hosted the workshop *Managing Protection Strategies: Measurability, Adaptability, and Evaluability of Protection* in London. Participants included a mix of protection and evaluation specialists from ALNAP and InterAction member agencies. In April 2017, ALNAP led a discussion on the pilot *Protection Guide* at the Danish Refugee Council's(DRC) Monitoring Evaluation and Learning Network meeting in Athens. Participants included protection and evaluation specialists from DRC country offices. In July 2017, ALNAP provided a training on protection and evaluation in AECID headquarters in Madrid. The workshop was attended by AECID humanitarian desk officers and evaluation specialists.
- 3 The ALNAP EHA Guide is the key ALNAP reference text on humanitarian evaluation covering, from a user-focused perspective, all the steps in an EHA from initiating the evaluation to communicating its finding and supporting its use.
- 4 In the context of protection work, issues of risk are framed using the Risk Equation: $\text{Risk} = \text{Threat} \times \text{Vulnerability} / \text{Capacity}$. See, among others, ICRC (2018: 39). See also Section 7 on complexity-sensitive evaluation methods.

Notes

A / Initiating the evaluation



A / Initiating the evaluation of protection

Content of this module at a glance

Section	What this section does	Who this section is for
4	clarifies the protection-specific evaluability conditions and opportunities to promote utility	evaluators and evaluation offices
5	suggests a framework for selecting evaluation questions linked to the intervention logic	primarily evaluation office and commissioning staff, but also useful for evaluators during inception phases
6	considers issues related to selection of indicators	evaluators and evaluation offices
7	provides guidance in the selection of approaches, designs and methods	primarily evaluation teams, but also for offices/evaluation commissioning staff when drafting ToR and assessing inception reports

'Initiating' an evaluation here refers to the different actions that should be undertaken before the data-collection work has begun. These can cover the initial scoping phase, the definition of the ToR, and the inception phase work of the evaluation team. The proportion of tasks undertaken before the evaluation formally begins, versus those undertaken during the inception phase, is likely to vary across organisations and assignments.

The considerations below are relevant regardless of the choice of evaluation type: ex-post or mid-term, formative or summative, real-time evaluation (RTE) or after-action review (AAR) (see [Section 2.2 of the EHA Guide](#) for guidance on the full range of evaluative options open to commissioning agencies). Indeed, many of the considerations outlined in this section are equally applicable to meta-evaluations and syntheses (e.g. those concerning the design of evaluation questions).

Investing in pre-evaluation and inception processes has emerged as an area of good practice¹ that can make an evaluation:

- better understood and more easily accepted and 'owned' by its primary intended users
- more useful to its ultimate users
- better supported by programme staff and championed by the leaders and managers who should take action on the conclusions and recommendations.

This first module should be seen as a reminder that improving the overall quality, usefulness and credibility of evaluating protection can start with:

- **Ensuring that evaluators are critically reflecting** on the systemic and organisation-specific features of protection in the context of humanitarian action.
- **Clarifying the protection-specific evaluability conditions** to take more informed and better-timed decisions on initiating an evaluation or reflective exercise.
- **Giving space for the evaluation team** to use the inception phase to build consensus on evaluation objectives and focus, thereby reinforcing ownership and opportunities to maximise utility.
- **Ensuring that approach, design and methods** are suited to the evaluation questions, expectations and field conditions facing the evaluation team.

4 Protection-specific evaluability conditions and opportunities to promote utility

Why do evaluability and utility analysis matter?

Evaluability assessments are important for at least two reasons. First, they can help inform a decision on whether or not to undertake an evaluation. If it is concluded that the evaluation should go ahead, the work done as part of the evaluability assessment can directly inform the evaluation ToR.

Second – when done well – evaluability assessments can help identify steps in undertaking the evaluation that will facilitate the process (UNEG, 2011: 17). For example, a good evaluability assessment will identify whether the key stakeholders have implicitly different understandings of what the evaluation should achieve, or even what the programme being evaluated is intended to change. Such issues can then usefully be prioritised during the evaluation inception phase. Likewise, a thorough evaluability assessment will identify any relevant gaps in the availability of data, or logistical and ethical challenges to data collection, that the evaluators will need to address in designing their methodology.

ALNAP's EHA Guide outlines the key considerations for undertaking an evaluability assessment that are applicable for humanitarian action as a whole. These include:

1. **Overall level of ambition and type of questions** that evaluation and programme stakeholders would like the evaluation to answer
2. **Programme design and intervention logic**, which is particularly important for outcome and impact evaluations that use theory-based designs to understand causation, and for mixed-methods designs and outcome-based approaches that look at the contribution to results in multi-actor or networked interventions (e.g. outcome mapping; outcome harvesting; RAPID Outcome Mapping Approach – ROMA)
3. **Availability of data** or feasibility of generating data with the resources allocated, so that the evaluation can answer the chosen questions.

4. **Conduciveness of the context** to carrying out an evaluation, for example in terms of access, logistics and security, and also the local office's ability to host the evaluation team, as well as the organisational 'climate' and leadership support for the evaluation.

For more information on how to address each of these points, and on how to go about conducting an evaluability assessment, see [ALNAP EHA Guide Section 2.4](#).

Protection-specific evaluability conditions

Protection evaluations present their own specific constraints to evaluability. These constraints can apply regardless of whether the evaluation is looking at protection-specific activities; the integration of protection concerns in sector-specific projects; or efforts to mainstream protection across all humanitarian activity. As a result, evaluability assessments for protection evaluations should try to address the extent to which these conditions hold.

[Toolkit item # 1](#) provides a checklist of questions to ask in an evaluability assessment for a protection evaluation. Evaluation commissioners are encouraged to use this toolkit as the basis of an evaluability assessment wherever protection concerns are an important area for investigation in the evaluation.

Critical issues to consider include:

1. **Overall level of ambition and type of questions:**

- Is the evaluation expected to look at results that are beyond the control of the agency's intervention?

Although not overtly relevant to protection, this question is particularly important because of the significant role played by external actors in achieving protection outcomes and in establishing and maintaining the protection environment.

- Is the evaluation expected to reach across multiple domains in the protection architecture, including development aspects?

Here again, this question often applies beyond protection evaluations, but is particularly important because of the broad ways in which humanitarian agencies define protection work. Some interventions aim to improve protection outcomes by improving household resilience, for example. In such cases, it is important to clearly define the scope of the evaluation from the outset.

- Are there realistic expectations for the evaluation?

Given the issues highlighted above, it is especially important to match expectations to resources when undertaking protection evaluations. The desire to look at higher-level results and environment-wide issues needs to be tempered by the available resources, time, evaluation expertise and protection know-how.

2. Programme design and intervention logic:

- Does the programme clearly define the problem it aims to address? Is the expected change related to protection?
- Are the drivers of protection risks identified in the assessment, programme documents, or result framework?
- Is the programme's results framework coherently articulated? Do the outputs, outcomes and goal follow a coherent logic? How does protection feature in the resulting framework? For example, is it viewed as a stand-alone activity? As a specific activity integrated alongside other sectoral activities? Or is it understood as an issue to be mainstreamed across all project activities?

3. Availability of data:

- Has the programme or intervention generated the data needed to carry out disaggregated analysis by sex and age (at minimum), and by other characteristics, vulnerabilities, or other lines of affiliation to groups and sub-groups depending on the context and programme being evaluated?
- Was the initial programme or intervention design based on a protection analysis or other type of assessment or baseline study? If so, was disaggregated data collected for that analysis?

4. Conduciveness of the context:

- Would the internal conditions of the programme/project and the broader external conditions of the context within which it is situated allow for an evaluation to take place? Are conditions conducive to ethical primary data collection and field visits?
- Are there resources, timing and security restrictions that should be taken into account at the scoping and design stage of the evaluation?

5 Evaluation questions and the intervention logic

Evaluation questions

Evaluation questions frame the focus of the evaluation and can help to tell a comprehensive story when the findings are presented (Kuster, et al., 2011: 40). ALNAP's EHA Guide provides a detailed breakdown of the role that questions play in an evaluation of humanitarian action and how best to select them. For more information on this topic, please refer directly to the [EHA Guide, Section 6](#).

One critical issue raised in the EHA Guide is the relation between evaluation questions and evaluation criteria. In evaluations of aid interventions, the most common framework for structuring evaluation questions are the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) criteria, which focus on relevance, effectiveness, impact, efficiency and sustainability (OECD-DAC, 1991). These have been adapted to focus on issues arising in humanitarian action to comprise coverage/sufficiency, effectiveness, relevance/appropriateness, efficiency, connectedness, coherence and impact (OECD-DAC, 1999).

As noted in the ALNAP EHA Guide, however, it is important that evaluation questions meet the information needs of the primary intended users, rather than simply responding to a set of pre-defined criteria:

At the EHA planning stage, first identify what you need to know and then place it within the evaluation criteria – not the other way around. The criteria are tools to think with and may suggest additional relevant questions. Ask evaluation questions only if you are ready to take action based on the answers, and use only the criteria that relate to the questions you want answered. It is the questions that matter, not the criteria. (ALNAP, 2016, p.112)

Thus, evaluation questions should be tailored to the specific needs and areas of responsibility of the evaluation stakeholders. Evaluation criteria are useful tools for this process, but they should not be used as a standard shopping list of questions for every evaluation to answer. This is particularly relevant in the case of protection, where different agencies will place different emphasis on distinct aspects of protection actions.

With this point in mind, this sub-section provides protection-specific considerations related to each of the adapted OECD-DAC criteria. These points should be taken into account when designing questions for protection evaluations, but not to create a definitive list of protection evaluation questions.

Coverage/sufficiency: Are the volume and distribution of resources sufficient to meet needs? To what extent are needs covered?

Humanitarian reviews often point out that in complex emergencies protection is the greatest need but that it receives far fewer resources than more visible, measurable and straightforward responses (SOHS, 2015). Evaluating protection in humanitarian action is an important part of mapping coverage and sufficiency. Questions might focus on the specific operational environment of a given agency, or the 'big picture' of the extent to which protection efforts meet overall needs. The latter may include some critique of biases within the humanitarian system, in which protection risks that are difficult to measure and address are given proportionally less attention than more straightforward relief assistance.

**Effectiveness: How well were humanitarian objectives met?
Was the response timely?**

In general, the primary focus of aid evaluations is **effectiveness**, often framed as 'results', which may also encompass 'impact' (see below). The starting point for effectiveness analysis in the case of protection has to be the stakeholder's understanding of protection outcomes. ICRC (2018) provides a definition based on the Risk Equation discussed in Section 3, above (namely, Threat x Vulnerability / Capacity). Unpacking the measurement of risk and relating it to the perspectives of each of the key stakeholder groups is a critical task for any evaluation of protection.

Given the frequently large gap between needs and operational capacities, and the pressures on agencies to promise ambitious results, assessments of effectiveness may also include a basic 'reality check' on the extent to which planned objectives were feasible. Further, the volatile context of humanitarian action in general and protection in particular may lead to a mismatch between rigid results frameworks and operational realities. Part of the 'reality check' is about learning how to adapt results to changing needs and operational opportunities. It is also important for evaluators to understand how to approach and measure protection outcomes.

Relevance/appropriateness: Do interventions address recipients' priority needs? To what extent do these needs drive programme design?

Ideally, evaluators should be expected to evaluate relevance against a pre-programme needs assessment and any further assessments undertaken throughout the initiative. Evaluators may face the challenge of whether and how to make up for these assessments' shortcomings. Resources generally allow (at best) for a review of secondary sources of information about needs and seldom allow an evaluation team to gather empirical data directly. Such data may be essential in the case of an impact evaluation, with obvious implications for whether such an evaluation is viable.

In addition to the relatively little attention paid to protection, noted above, perhaps the other great gap in humanitarian performance relates to engaging recipients in defining their own needs and programme design. The importance and contributions that this can provide are clear, as are the ethical imperatives. There are many methods for doing this (see [Toolkit item # 3](#)). There are also ethical challenges and dangers, discussed in [Module B](#).

Efficiency: Do outputs reflect the most rational and economic use of inputs?

While a seemingly straightforward criterion for evaluation, efficiency can be one of the most difficult to tackle in relation to protection. This is primarily due to the difficulty of quantifying protection activities and outputs. The evaluation of the ECHO response to the Syrian Crisis 2012-2014, for instance, reported that one of the main activities ECHO field staff undertook regarding protection was a form of continuous dialogue with partners, which supported the mainstreaming of protection across the programme portfolio. But the evaluation was not able to quantify the time and resources used for this activity, which hampered an efficiency assessment of this activity.

The picture can be further complicated if the commissioning agency does not provide a clear definition of what it wants to evaluate under this criterion. Efficiency is conventionally defined in terms of the ratio between inputs and outputs (see, for example, Beck, 2006). But some evaluations include other concepts under this criterion, such as cost-effectiveness and value-for-money, which reach beyond efficiency as such to include input–outcome and input–impact relationships, as well as other aspects (see, for example, DFID, 2011).

In the case of protection, it is especially challenging to assess input–outcome ratios, because of difficulty of quantifying protection outcomes. Recent literature has sought to explore the concepts of humanitarian and development applications of cost–benefit analysis, cost-effectiveness and value-for-money (e.g. Pongracz et al., 2016; Wheatley and Pongracz, 2015; BOND, 2011). Much of this literature addresses the challenge of providing alternatives to quantified cost–benefit analyses. Evaluators are advised to review the literature before selecting efficiency and/or cost-effectiveness metrics that are well suited to the specific nature of the protection activities and results being evaluated.

Connectedness: Do humanitarian interventions take account of other key actors and efforts?

The humanitarian sector is coming under increasing pressure to strengthen links to other key actors and efforts, humanitarian and otherwise. Questions in this regard may refer to two broad categories. The first is coordination within the system. The IASC (2016) Policy on Protection in Humanitarian Action provides a useful reference point for understanding the definitions, mandates and roles of different actors within the system regarding protection. But given the complexity of the coordination challenge, it is crucial for evaluations to understand the way that a given agency frames the concept of protection and strives to ‘connect’ with other actors and sectors.

Furthermore, considerable contextual analysis is required to understand which connections are appropriate, feasible and desirable. Ideally, the agency will already have undertaken this analysis, but this may not always be the case. (Re)constructing this analysis may be an important part of describing and critiquing an intervention’s logic. This may include asking what the initial assumptions were about who would do what, and whether those assumptions were valid.

Central to this, protection issues are often very closely associated to power and how it is used in society – at the level of the state, the community and the household. Interventions in this area inevitably become part of complex social processes which involve a range of actors. To be effective, the agency needs to be able to ‘connect’ to a complex social context, and an understanding of this context is often a precondition for making any judgement about the value of the programme.

The other aspect of connectedness concerns the extent to which an initiative relates to the broader protection environment. Any intervention should be designed to take into account the role of the state in protecting its population and respecting the rights of displaced populations. The extent to which an agency has the mandate, opportunity or ambition to enhance the role of the state or other national partners will vary. Here again, it is essential to explore this when initiating the evaluation.

Coherence: Does the intervention adhere to core humanitarian principles and align with broader peace and development goals?

In many evaluations the analysis of connectedness will overlap with that of coherence. Reflections on ‘who does what’ will inevitably need to be anchored in an understanding of ‘why they do it’. It may be assumed that there is broad consensus on core humanitarian principles and peace and development goals, but the interpretations may vary, as do the mandates and areas of engagement of different agencies.

Impact:

The definition of impact in humanitarian response is often contested and sometimes muddled with other criteria (SOHS, 2015). ALNAP’s EHA Guide provides an overview of the challenges to impact evaluation in humanitarian contexts. Briefly, these include:

- **The attribution challenge:** impact evaluations seek to provide causal statements regarding the effects of programme interventions. But complex and fluid humanitarian crises often make it difficult to attribute impact-level effects (such as improvements in food security at household level) to programme activities (such as provision of cash assistance to food insecure populations). As outlined by the EHA Guide, this is primarily due to the significant role of external actors in complex environments such as those faced by humanitarian actors, and the unpredictability of causal chains. Arguably, it is also harder to establish comparability between treatment and control groups in humanitarian contexts. See [Section 12 of this Guide](#) for further detail.

- **Data availability:** baseline data on outcomes (such as food security levels) is often missing in fragile and conflict affected states prior to crisis onset. This makes statistical tests such as difference-in-difference difficult to use. Demographic data on populations in need of assistance can also be missing.
- **The need for rapid action:** humanitarian crises often demand rapid response mobilisation, which can present challenges for the planning of impact evaluations (ALNAP, 2016, pp.357-358).

Section 18 of the EHA Guide provides an overview of these challenges and suggests some potential solutions. Nevertheless, several of these issues – most notably attribution – are exacerbated in crises that have significant protection elements. As a result, evaluators and evaluation commissioners should think carefully about how to address impact in protection evaluations. It is of paramount importance to avoid making false or superficial attribution claims.

Intervention logic and theory of change

As with all evaluations of humanitarian action, it is important for protection-related evaluations to unpack and interrogate the intervention logic or theory of change (ToC). Section 5 of the EHA Guide presents an overview of the different types of intervention logics used in the humanitarian sector, as well as tips for evaluators on how to unpack them.

Protection-specific evaluations present particular challenges to this endeavour, arising directly from the nature of protection actions. Section 1 of this Guide notes that:

- Protection programme models can take many different forms: stand-alone, integrated or mainstreamed. Confusion about the programme model can make it harder to conduct a good evaluation.
- Protection outcomes are often dependent upon a wide range of external actors and factors that may contribute towards – or hinder – the creation and maintenance of a positive protection environment.

- Protection results are difficult to predict. Humanitarian emergencies are complex systems, with many interconnected parts linked by hard-to-predict causal pathways. Protecting affected populations means reducing a range of risk factors (physical, legal, economic) that can be spread right across such a complex system. This can make protection results very hard to predict at the programme start.

For these reasons it is essential that evaluation commissioners and evaluators take a careful approach to interrogating the initial programme ToC, with particular attention to the following objectives:

- Determine which programme model best fits the intervention: stand-alone, integrated or mainstreaming.
- Understand how the intervention fits with the policy, institutional and conflict context .
- Keep an open mind about potential unintended consequences of the intervention.

Specify the programme model:

It is always vital for evaluation stakeholders to reflect on the shape of the programme model and how it relates to protection risks for the affected population. Although this work should normally be covered by programme staff during the design phase, it remains important that evaluation staff (commissioners and evaluators) critically consider the programme model themselves before designing the evaluation. Doing so should encourage a tailored evaluation approach and reduce false assumptions about how (and if) the programme intended to reduce protection risks.

The following examples represent **three illustrative cases** of how protection can relate to a programme. They are intended as a tool for reflection and do not constitute a typology of protection programming or point towards a set of good practices.

The examples illustrate different ways in which humanitarian actors address and weave protection into interventions. The order of the three examples is in no way intended to suggest judgements on their relative appropriateness.

Example 1 | A programme where protection is to be achieved through specialised or dedicated actions

Evaluators may be asked to evaluate a project or programme where protection is achieved through specialised or dedicated actions.

These programmes are often described in the literature as vertical or stand-alone protection actions.

They are often characterised as the traditional remit of protection actors with a specific mandate anchored in international legal instruments and of other actors (including NGOs and international NGOs) with specific expertise in thematic areas – such as forced displacement, child protection, or working with people with disabilities, the elderly and sexual and gender minorities.

Programme features from a protection angle

The desired outcomes of these interventions explicitly articulate and speak to protection issues (Davies and Ngendakuriyo, 2009; de Sas Kropiwnicki, 2012). There are expectations that protection actions in this example are:

- informed by a protection-specific mandate, policy and/or strategy, and that a protection risk analysis (using, for example, something akin to the Risk Equation presented in ICRC (2018) and elsewhere) is explicitly used to inform the protection strategies and logic of the intervention/programme
- explored in the conflict analysis that should generally underpin and inform humanitarian response strategies
- anchored in different modalities and lines of work connected to protection (as discussed in the Introduction section).

Murray and Landry (2013: 5) note that protection actions in this example are usually featured in the 'protection chapter' of some key humanitarian funding tools such as the consolidated appeal process.

Some protection actions of this type may aim to influence outcome-level changes in the broader protection environment.

Protection actions such as these are likely to be **highly sensitive to agencies' mandates**. There are higher expectations that agencies with a specific protection mandate have greater capabilities to raise funds, design, carry out, lead and coordinate interventions and dedicated protection-oriented programmes and services of this type.

Example 2 | A programme where protection is integrated into other sectoral and multi-sectoral interventions

An evaluator might be asked to consider a project or programme in which protection goals are to be achieved or supported by integrating protection-oriented activities into other sectoral and multi-sectoral programming. This implies applying protection-related perspectives and activities within an intervention such as water and sanitation, education, health, food security, livelihoods, or shelter.

Expected programme features from a protection angle

Protection actions in this example are unlikely to be explicitly anchored in a given **agency's mandate**. Services and actions oriented towards addressing specific protection risks or reducing exposure to protection risks are integrated into other ongoing sectoral and multi-sectoral interventions. As a result, protection-related objectives may be vaguely formulated and/or based on a relatively weak analysis of the overall protection context. On the other hand, a sectoral perspective can also reveal new protection risks and challenges that are not apparent to 'protection experts', whose frames of reference are more focused on 'conventional' protection concerns.

Protection actions such as these are also likely to encompass a variety of implementing approaches – from information provision to operational advocacy and provision of specialist services.

Make space for the policy, institutional and conflict context:

One good example of a reconstructed intervention logic that accounted for the policy, institutional and conflict contexts was demonstrated in the pilot phase of this Guide. The Danish Refugee Council (DRC) evaluated its humanitarian response for Syrian refugees in Southern Turkey in 2015 (DRC, 2017). The response included a mix of in-kind and cash-based transfers to meet the basic needs of new arrivals while also aiming to reduce protection risks for populations affected by protracted displacement. The evaluators sought first to unpack the ToC and identify the outcome-level objectives of the programme. But in doing this, they also identified a significant change in the policy and institutional context arising from the establishment of the Emergency Social Safety Net, which sought to provide the most vulnerable refugees with a social safety net to meet their basic needs (see, among others, ECHO, 2018).

Example 3 | When protection is mainstreamed across all programming

An evaluator might be asked to consider a project or programme in which protection goals are to be achieved or supported by integrating protection-oriented activities into other sectoral and multi-sectoral programming. This implies applying protection-related perspectives and activities within an intervention such as water and sanitation, education, health, food security, livelihoods, or shelter.

Expected programme features from a protection angle

Some of the main points arising from this example are:

- Protection actions are not always a matter of protection-specific mandates and they are not necessarily undertaken by protection specialists
- Programme documents might not include any direct reference to 'protection', even though they may involve significant protection goals
- There may be references to agency guidelines, agency-specific commitments relating to protection in humanitarian action, but not to how these should be translated into programming actions – including in resource allocation – in specific sectors
- Protection may even be overlooked entirely in the intervention logic, which in turn is likely to place the onus of reconstructing the protection-specific intervention logic on the evaluation team.

*** Note on the term 'safe programming':** Safe programming refers to any attribute and way of working of the programme/service or other type of intervention that aims to ensure that interventions: (a) do not put the population in danger; (b) contribute to their security as much as possible; and (c) analyse and monitor potential threats in a systematic way.

Source: Oxfam GB (2009)

As the evaluators put it:

The ESSN is coming to dominate the overall framework for humanitarian response, and with that protection. The extent to which the ESSN reaches the most vulnerable and is tailored to overcoming the risks that these people face will impact on the roles of the broader humanitarian community in mounting remedial operations. The environment for protection is shifting from a primarily humanitarian protection context to one where humanitarians are focusing their efforts on highlighting and addressing gaps in the social protection environment as refugee needs are increasingly addressed within [Government of Turkey] systems. (DRC, 2017: 8-9, original emphasis)

As a result of this analysis, the evaluators were able to disaggregate the elements of DRC's response that were under the programme's direct control, those that lay within its sphere of influence, and those that lay outside this, but inside its sphere of interest:²

Figure 3: Spheres of interest, influence and control

Sphere of interest	Sphere of influence	Sphere of control
<ul style="list-style-type: none"> • labor market for refugees (and local population) • conflict dynamics • overall government policies 	<ul style="list-style-type: none"> • awareness of protection issues • adaptation of 'good practice' in CBIs • beneficiary and authority awareness of rights to basic services • creation of referral systems and access to Turkish government health services 	<ul style="list-style-type: none"> • winterization protection • vulnerable populations' access to basic supplies and services • financial resources for food, etc. • identification of highly vulnerable households to receive targeted assistance • links to services from other international agencies

Source: Danish Refugee Council (2017)

This both provided a clearer understanding of the role of external factors in the programme than was previously available, and helped to frame evaluation expectations by identifying areas where contribution analyses could and could not be meaningfully applied.

Consider unintended consequences:

Evaluators and evaluation commissioners need to be careful when using the ToC to guide the evaluation of a programme. Given the complexity of protection interventions, it will always remain difficult to understand their full range of consequences. Evaluations should be able to interrogate both the intended and unintended effects of any given programme. It is therefore important to allow space for open-ended enquiry during the evaluation process, rather than simply seeking to establish whether or not the original intended results were achieved.

In this regard, recent work led by InterAction on the Results-Based Protection framework has explored the applicability of systems-thinking to protection programming, including by making space for adaptation to changing risk patterns over the course of the programme (see, for example, Acaroglu, 2017). This has important implications for evaluators. It makes it vital that evaluations seek to understand how the programme adapted and evolved in response to changing protection risk profiles during the programme cycle.

Tools such as Outcome Mapping are important here, and [Toolkit Item #4](#) presents an approach to developing theories of change that can provide a useful reference point for evaluators seeking to explore how a programme has led to behavioural changes.

Other tools drawn directly from complexity science can also be used to help capture important elements of a programme and its effects that might otherwise be missed. Bamberger et al. (2016) provides a useful overview of such approaches, including:

- System dynamics: which can help to map the programme context and test underlying assumptions in the theory of change
- Critical systems heuristics: which can help unpack the different norms, beliefs, values and perspectives held by the different programme stakeholders
- Systems mapping: which can help provide a comparison between the wider context in which a programme sits, before-and-after programme completion
- Social network analysis: which can help explore the layers of interrelationships between different programme stakeholders in order to better understand and trace programme results
- Agent-based modelling: which can help evaluations to explore complex causal processes through the simulation of interactions, preferences and characteristics of individual agents (Bamberger et al., 2016: 39)

We return to the issues surrounding intervention logics in [Module C on analysis](#).

6 Indicators

One of the most challenging aspects of planning an evaluation is to select rubrics (see [Toolkit item #2](#)) and indicators that are relevant for the evaluation questions. Ideally, these will have been identified in the intervention's results framework (at least regarding effectiveness) and relevant data monitored. In practice, however, initial results frameworks may not correspond to the issues of interest in the evaluation, or they may have been rudimentary and/or ignored in programme implementation.

Moreover, as discussed in Section 7 below, good protection programming can entail the use of continuous context-specific protection risk analysis, and subsequent programme adaptations as a result of a changing environment. In these cases, initial programme indicators can become less relevant and there may be a greater need for new ones. Taken to the extreme, the developmental evaluation approach outlined in Section 7, below, encourages evaluators to work alongside programme units through the adaption process, implying that evaluation indicators themselves will evolve during the evaluation cycle.

However the process unfolds, protection evaluations require indicators that measure the quality and value of both outputs and outcomes. This is likely to include an analysis of quality and value of the intervention as perceived by the participants, programme recipients and other key stakeholders. Examples of indicators for protection evaluations can be found in, among others, the UNOCHA Indicators Registry or UNICEF (2016a).

Protection-specific challenges for developing evaluation indicators include:

- **Data availability** can often be especially difficult in conflict-zones where protection risks are prevalent. Recent research by the SAVE consortium provides examples of tools and approaches to overcome the challenge of monitoring in conflict environments (Steets, Sagmeister and Ruppert, 2016). From an evaluation perspective, the evaluation assessment should outline the underlying assumptions on the feasibility of accessing the data and the expected level of confidence in them. If necessary, alternative indicators may need to be selected. Some might be proxy indicators, as discussed in the following section.

- **Defining indicators for integrated protection activities** presents specific challenges. Programmes that seek to integrate protection activities in other sectoral projects can be especially challenging in terms of defining indicators of success. It is important to ensure that both protection and sector-specific indicators are included in the evaluation matrix. ECHO (2016: 57) provides some useful considerations in this regard, including the suggested use of the Coping Strategies Index as a way of tracking cross-sectoral outcomes.
- **Defining indicators for protection advocacy.** The evaluation of advocacy efforts is both difficult and essential to protection evaluation. Section 13, below, outlines some of the key challenges and approaches, although as yet there is little humanitarian-specific guidance on the challenges of evaluating advocacy. One recent example of good practice in a protection-specific context is NRC's evaluation of its protection and advocacy work in the DRC for the period 2012-2013 (O'Neil and Goldschmid, 2014).

Proxy indicators

Proxy indicators track measurable changes that are understood to represent the occurrence of a related but unmeasurable change. Proxy indicators allow evaluators to assess the likelihood that a change has occurred even when direct measurement isn't possible, practical or preferable (Corlazzoli and White, 2013: 20-21).

In EHA in general, and in evaluating protection in particular, proxy indicators are helpful in situations where standard data-collection mechanisms are insufficient. This may be because monitoring mechanisms have been disrupted or because certain lines of enquiry for primary data collection are not feasible or appropriate (or even ethically defensible, as discussed in Module B) (also in ICRC (2018), Chapter 2).

Proxy indicators also offer a way to measure more abstract concepts, such as well-being, trust or community cohesion. When looking at these aspects of performance, such indicators will be important in bringing contextual factors into focus, but there are certain pitfalls.



- When proxy indicators make use of secondary data – such as administrative data and data logs from helpline, injury surveillance and health centres, for instance – they are prone to **bias from usage**: they only capture cases that have been reported or detected or for which services were sought.
- It is important to understand, and be explicit about, the assumptions that underlie the belief that the proxy is indicative of the desired result. For example, a programme might aim to increase access to psychosocial care for in-camp refugees, for which an evaluation team might use a proxy indicator such as the number of new psychosocial care units opened inside a refugee camp. The underlying assumption is that access to psychosocial care is primarily limited by the supply of care units rather than other factors. Clarifying these assumptions can improve the quality of the analysis drawn from the indicator.

Good practice examples: Using administrative data as proxy indicators

Data from health centres is potentially useful for triangulation or as a proxy indicator, especially if they cover a particular response linked to child protection. For example, the 2009 Kenya situation analysis used data from the Gender Violence Recovery Centre of the Nairobi Women's Hospital, and triangulated it with survey data, caseload reports and official reports of the Kenya Police to establish changes in violence over time.

Data from child helplines can be useful, for example, to triangulate information from other sources like surveys and police data. In the absence of any other data, they could also be reviewed for a basic, highly aggregated analysis and/or to establish trends over time.

Source: UNICEF (2015: 46)



Good practice example: Finding indicators for hard-to-measure issues

In a collaborative approach to psychosocial programming, a number of Palestinian agencies agreed to specific indicators of aspects of psychosocial well-being: for example, reduction in troubling dreams (as a measure of emotional well-being) and increasing collaborative behaviour with teachers and peers (as a measure of social well-being).

Identifying such indicators has enabled psychosocial workers to gather clear results on their interventions. It has also led to more awareness of these aspects of behaviour among teachers, and to an increase in parental involvement in children's activities at school.

Source: IFRC (2009: 175) cited in Ager et al. (2011)

Proxy indicators are particularly important when the evaluation involves measuring an impact in terms of things that did not occur, for example human rights violations or incidents of GBV. For example, measuring reductions in GBV in general may provide a proxy measure for assessing efforts to strengthen the protection environment and the role of the state. An evaluation can measure relevant trends as a proxy indicator, while recognising that these trends can seldom be directly attributed to the intervention.

In the sectors of conflict transformation and security and justice, one area of emerging evaluation good practice concerns the use of bundles (or baskets) of proxy indicators to help measure broader trends by looking at nuances of change (Corlazzoli and White, 2013). For example:

- In Bangladesh, hospital admission records have been used to verify media reports of acid attacks against women.
- In Afghanistan, parental perceptions of safety (garnered through FGDs) were combined with school attendance data to determine trends in freedom of movement.

As highlighted in the Afghanistan example, data from another sector can be useful. School attendance is increasingly used as an indicator of freedom of movement. In the same vein, indicators relating to patterns of participation in the public sphere (such as cultural, religious events, or weekly markets) are also increasingly used as part of the 'basket of indicators' on freedom of movement.

Of course, all such approaches need to bear in mind the potential risks of identifying individuals through the aggregation of data sets, and the related legal implications on data protection and privacy risks to the individuals concerned. [Module B](#) discusses these issues in further detail.



Evaluator's insight: Thinking outside the box to gather data

Creativity and imagination can generate new data-collection techniques for evaluative analysis.

For example, an evaluation that sought to determine the protective benefits of World Food Programme food relief in Darfur used proportional piling of beans to understand sources of livelihoods and whether people had to take risks by leaving displacement camps to seek work or gather firewood. Interactive theatre can also be used as an evaluative measure, with the drama stopping at critical junctures and the audience actively choosing how a story should proceed, and recording those decisions. This audience input is a public statement, however, and so caution is needed to ensure that people taking a public stand are not put at risk.



Keep in mind: Using official data for protection M&E

There are two very different sources of official data used by conflict crime and violence programming that are relevant to protection programming monitoring and evaluation (M&E) – data sets based on criminal justice and on public health:

Criminal justice sector – collects primary data categorised as crime in the respective legal system. Recording can occur in various places – e.g. recording a homicide could occur (a) as a body is found (police data), (b) as it is autopsied (forensic data) or (c) as criminal prosecution ends in a judgement (ministry of justice data).

Public health sector – collects primary data on violent deaths as they occur in hospitals or health care (e.g. intensive care units), or as deaths are recorded in national vital registration statistics (usually under the ministry of health).

The key difference is the focus on **events** (e.g. police records of the number of crimes) or on the victims (e.g. emergency room services), although the distinction is not absolute. The availability and quality of the data vary widely, and some argue that homicide data is the strongest, because as a very serious offence it is more likely to be recorded in criminal justice records. Conversely, public health data may have a better chance of picking up on the scale and effects of violence (except in the case of mental health, as provision tends to be poor).

Some secondary data sources, such as observatories (or Armed Violence Monitoring Systems) combine both criminal justice and public health sector data. The Jamaica Crime Observatory, for instance, maps data from the Jamaica Injury Surveillance System onto police crime data. State capacity to compile data sets in both criminal justice and public health may, however, be weak.

Source: Small Arms Survey (2013)

7 Selecting approaches, designs and methods

Every evaluation requires an approach, design and methods that are: (1) in line with its purpose and questions; and (2) responsive to the features of the programme, the specific programme components and sub-components being examined, and data and contextual factors. The ALNAP EHA Guide provides a detailed presentation of methods applicable to humanitarian action, and considerations regarding their selection.

Protection work, however, presents a number of issues that challenge evaluation methodologies in specific ways. These include the:

- nested nature of many protection actions within other programmes and presence of different strands of work that co-exist in protection in humanitarian action
- likely presence of spill-overs between different lines and modalities of protection work
- complexity of protection actions, which, as discussed above, often entail high degrees of interconnectedness and complex causal pathways
- heightened requirements for legal, ethical and conflict sensitivity considerations in the evaluation process, particularly in the data-collection stage (discussed in detail in ICRC (2018, Chapter 6)
- importance of advocacy aimed at duty bearers as an important modality for protection actions, which in turn presents evaluators with all of the challenges that evaluating advocacy has always faced.

Specific considerations regarding the challenges of evaluating advocacy are presented in [Section 13](#) and issues regarding data collection and use in conflict contexts form the basis of [Module B](#). This section presents two considerations that evaluators should take into account when selecting evaluation methods in order to overcome the first three issues above, namely:

- mixed-method approaches
- complexity-sensitive methodologies.

Mixed-method approaches:

The first consideration concerns the importance of taking a mixed-method approach. Evaluative thinking outside the humanitarian sector has moved away from preconceptions about a single 'gold standard' and a presumed need for quantitative impact evaluations. Patton (2014) suggests a need to 'up the ante and aim to supplant the gold standard with a new platinum standard: methodological pluralism and appropriateness'. The goal of Patton's 'platinum standard' is to find the most appropriate blend or best fit of designs and methods to answer the evaluation questions at hand (Alexander and Bonino, 2015: 13-14).

This point is particularly pressing in cases like protection, where programmes often engage a range of different stakeholders in different ways. This may be due to the nested and overlapping nature of protection programme models, as mentioned above, or to the nature of protection itself, where victims, aggressors and duty bearers can have very different perspectives, incentives and interests.

In such a situation, it is wise to admit that no single evaluation approach can cover all needs, and therefore combine a variety of methods to increase the credibility of the findings.³

There are three main types of mixed-method design:

- **Sequential mixed method:** Quantitative method followed by a qualitative method or vice versa.
- **Parallel mixed method:** Quantitative and qualitative components conducted at the same time.
- **Multi-level mixed method:** Where a large evaluation is conducted at multiple levels, with both quantitative and qualitative approaches being used at each level, including analyses of both direct protection actions and also the protection environment.

Table 1: Techniques that can strengthen mixed-method design

Technique	How it bridges qualitative and quantitative methods
Triangulation	<p>The combination of data, analysis and findings generated from qualitative and quantitative analysis can increase the strength of the causal inference.</p> <p>In general, the validity of the evidence generated using triangulation is enhanced when two or more independent estimates can be compared.</p> <p>Both quantitative and qualitative methods use triangulation, but often in slightly different ways for example:</p> <ul style="list-style-type: none"> • Quantitative methods are usually more concerned with using triangulation to check the internal consistency of measurements (e.g. for survey instruments). • Qualitative methods prioritise the use of triangulation to verify the information gathered and deepen and broaden understanding of an issue or phenomenon through obtaining multiple perspectives from different sources. This is done, for example, when the use of purposive sampling requires rich explanations and narrative that can illuminate cases of interest.
Process tracing / Process analysis	<p>Qualitative analysis focused on processes of change within cases may uncover the causal mechanisms that underlie quantitative findings. Without process analysis it may not be possible to assess whether failure to achieve a certain outcome is due to the failure of design or implementation.</p> <p>Most useful as a method for identifying, testing and validating hypothesised causal mechanisms within case studies. Examination of multiple cases may be used to build up a body of evidence.</p>
Focus on tipping points	<p>Qualitative analysis can explain turning points and crucial junctures for change within quantitative time series and changes over time in causal patterns established with quantitative data.</p>
Using quantitative data as point of departure for qualitative research	<p>A quantitative data set can be used as a starting point for framing a study that is primarily qualitative.</p>

Sources: Tarrow (2009, reproduced in Stern et al., 2012: 33); Bamberger, Rugh and Mabry (2012: 229; 326-330); White and Phillips (2012)



Good practice example: UNICEF's use of a mixed evaluation method for its child protection response to the 2004 tsunami in Indonesia

The evaluation, which was commissioned by UNICEF's Child Protection Department, aimed to determine the impact of UNICEF's response to the South Asian tsunami within the child protection sector. It followed the evolution of the three strands of child protection work in the programme in Aceh (children without family care, children without psychosocial support and victims of exploitation and abuse).

The evaluation employed a sequential mixed methods approach to combine comprehensive coverage with in-depth analysis. It focused on three districts to compare results in mainly tsunami-affected and mainly conflict-affected districts, which allowed for comparisons between those areas with a strong operational UNICEF presence and those areas with less. The evaluation design also compared different interventions with one another or, where a similar programme did not exist, with groups of children who did not receive the intervention. For more, see www.unicef.org/evaldatabase/index_59604.html.

Source: UNICEF (2015: 201)

Further, different methods will usually be required depending on whether the design will be single level (e.g. the household, organisations or institutions) or multi-level (e.g. a country programme that requires description and analysis of links between different levels) (Peersman, 2014: 4).

Regardless of the specific mix or preference for a given set of data-collection and analysis tools, there are a number of techniques that can strengthen the evaluation design by bridging the use of qualitative and quantitative data-collection and analysis methods (see Table 1 on previous page).

Complexity-sensitive methodologies

As discussed earlier in Section 1, protection work is often a complex endeavour. The work of InterAction under the Results-Based Protection initiative highlights how protection programming and contexts constitute complex systems with dynamic, interconnected actors and parts. This increases the importance of adaptive approaches to management programming based on continuous context-specific protection analysis.

This point is echoed in the most recent update of the ICRC Professional Standards for Protection Work, which underlines the importance of basing programming on continuously updated risk analyses, remaining open to new information and unexpected consequences, and then making iterative course adjustments where necessary (ICRC (2018), Chapter 2).

Likewise, the recent Results-Based Protection work led by InterAction and ECHO's updated 2016 protection policy also emphasise the importance of continuous risk analysis. Arguably, this approach has its roots in previous risk-based approaches to protection as developed in Slim and Bonwick (2005).

Evaluations of protection need to be sensitive to these points. This is, in part, a question of including evaluation questions that cover the use of continuous protection analysis by the programme team, donor or implementing organisation. It also concerns the inclusion of questions that explore the rationale behind programme adaptations taken during implementation: were they based on observed changes in need or other concerns?

There are many useful starting points for evaluators looking to assess the appropriateness and use of context-specific protection risk analysis, and evaluation teams should be encouraged to tailor their approaches to each case. Examples of good practice include:

- ICRC (2018), Chapter 2, presents an overview of the key elements of protection risk analysis.
- InterAction (2017a; 2017c) provide similar overviews, and examples of tools that can feed into continuous risk analysis such as the UNHCR and DRC Protection Information Management initiative.⁴

- ECHO (2016) provides a toolkit for identifying threats, vulnerabilities and capacities.
- Evaluations such as UNHCR (2016) and ECHO (2017) (both of which covered the South Sudan crisis) show how evaluations can address the conduct of a protection-specific risk analysis, its impact on programme adaptations, and the challenges of conducting protection risk analyses in sensitive contexts where duty bearers are not always willing to openly discuss protection risks.

Beyond helping to frame the evaluation questions, the complexity of protection programmes and contexts calls for considering evaluation approaches that are well suited to make sense of complex situations where other, more well-established methods, often fail. Of course, options are always constrained by budgets, short timelines and the pressing need for rigorous evidence. There is, moreover, a lack of tried and tested complexity-sensitive methods in the humanitarian evaluation canon. Nevertheless, there are alternative approaches, methods and tools from outside the sector that can be considered when designing evaluations of protection. These include:

- **Developmental evaluation:** an evaluation approach designed by Michael Quinn-Patton that aims to support innovation and adaptation in complex environments. Developmental evaluation is iterative and seeks to develop metrics and tracking mechanisms as outcomes emerge, allowing metrics to change during the evaluation implementation as the process unfolds (see, most notably, Patton, 2010; 2006).
- **Outcome harvesting:** an approach whereby evaluators collect evidence of outcomes and then work backwards to determine whether and how an intervention has contributed to these changes. This approach is particularly useful in complex situations where it is difficult to provide concrete definitions of the programme's objectives from outset (Wilson-Grau and Britt, 2013).

- **SenseMaker®**: a data-collection and analysis method, with an accompanying analytic software tool. It is designed to monitor and explore complex processes from multiple human perspectives. It consists of several micro-narrative collection tools, a pattern detection software and an underlying methodology.⁵ Such approaches can help evaluators to maintain an open-ended approach to identifying programme outcomes, rather than relying on pre-defined expected outcomes.

In some cases, the use of methods and tools such as those above has the potential to merge the boundaries between evaluation and project monitoring. Although it is important to understand and respect the different objectives between monitoring and evaluation activities, it is also important to provide evaluators with a set of tools that can respond to the complexity of protection activities and contexts. On this basis, we encourage organisations to consider such tools when designing, commissioning and conducting protection evaluations.

Endnotes

- 1 See for, instance, Hallam and Bonino (2013) for a study specific to humanitarian evaluation practice, and Rist, Boily, and Martin (2011) and Heider (2011) for useful insights from broader development aid evaluation practice.
- 2 Although similar to the terminology used by Outcome Mapping methods, the categorisation used in DRC (2017) is actually somewhat different. Critically, the sphere of interest describes elements that are more relevant as background contextual factors than as features of the environment that the intervention seeks to change, even by indirect means.
- 3 This section draws on Stufflebeam and Coryn (2014); Patton (2010); Church and Rogers (2006); Stern et al. (2012); Bamberger et al. (2016); Bamberger, Rugh, Mabry (2012); Chigas, Church and Corlazzoli (2014); Rogers (2012); Tsui, Hearn and Young (2014).
- 4 See <http://pim.guide/>
- 5 See <https://senseguide.nl/what-is-sensemaker/>

Notes

Notes

B / Data management



B / Data management: legal and ethical implications

Content of this module at a glance

Section	What this section does	Who this section is for
8	provides guidance for how to ensure that evaluations are carried out in a lawful, protective and conflict-sensitive manner	primarily evaluation teams
9	delves specifically into how to select data sources and address constraints in data management	
10	explores how to approach data gathering on less tangible dimensions	

The choice of the most appropriate approaches to data collection, management and use for evaluations in humanitarian contexts has **legal and ethical implications**. The stakes are likely to be high due to the challenges and constraints described above, as well as to the heightened need to comply with data-protection laws and for ethical safeguards and protocols to inform data collection. Consequently, and in addition to the considerations presented in this Module, evaluators and evaluation commissioners should consult the updated ICRC Professional Standards on Protection Work (ICRC (2018), Chapter 6), particularly the chapter covering data and information management in protection contexts, and the Handbook on Data Protection in Humanitarian Action (Brussels Privacy Hub and ICRC, 2017).¹

This section starts with an overview of the challenges and constraints in managing data for evaluating protection in humanitarian action. It then presents some ethical and practical implications that should be considered when taking decisions about:

- which type of data can be collected for which purpose
- from which sources and from whom the data should be collected
- minimising possible negative and harmful repercussions relating to data management.

8 Ensuring that the evaluation is carried out in a protective and conflict-sensitive manner

When evaluators are asked to look at humanitarian protection, it is critical to apply a ‘protection perspective’ throughout the process. This is relevant to all evaluations carried out in conflict and insecure settings, or other situations of violence. The starting point – as with general EHA practices – should be a consideration of ‘do no harm’ regarding those taking part in the evaluation process, in particular during fieldwork.² (See ‘Keep in mind’ box opposite for more on the ‘Do no Harm’ concept.)

Moreover, sensitivity to conflict is appropriate not only for conflict settings.³ Evaluators should be aware of how any evaluation could contribute to exacerbate tensions and compromise future access to affected populations. For example:

- Poorly conceived evaluation approaches may introduce biases in data collected and lead to suspicions of partiality between stakeholder groups (it is important to respect the principles laid out in ICRC (2018), Chapters 1 and 6 on this point).
- Expectations may be raised that by taking part in the evaluation (e.g. during data collection) more aid will be provided, which could result in frustration on behalf of affected people.
- The views of individuals and groups that are most at risk may be exposed, making them more vulnerable to reprisals by powerful actors.



Keep in mind: What do we mean by ‘do no harm’

The meaning of the term ‘do no harm’ differs in the field of humanitarian action and conflict sensitivity. ‘Do no harm’ is derived from the principle that a medical practitioner should cause no harm to the patient. It has been adopted and adapted in other fields (see also ALNAP, 2016).

From a **humanitarian perspective**, ‘do no harm’ is a term that is widely used but often not well defined. Nevertheless, ICRC (2018) provides an important and clear application of the principle to the protection context. The primary concern, outlined in Section 1.4, is to avoid aggravating or generating ‘additional protection risks for populations at risk or subject to violations and abuse’ through poorly conceived or carelessly implemented protection activities. Protection actors are asked to keep in mind that protection activities can inadvertently stigmatise individuals or communities who may be seen as providing sensitive information.

From a **conflict-sensitivity perspective**, ‘Do No Harm’ (capitalised in this Guide) refers to a specific seven-step framework that can be used to assess the conflict sensitivity of an intervention. It was developed by Collaborative for Development Action (now CDA) and is the most widely used tool for assessing conflict sensitivity. Conflict sensitivity means ensuring that an intervention does not inadvertently contribute to conflict and, where possible, contributes to peace (within the confines of an organisation’s mandate). In this definition, ‘Do No Harm’ relates to conflict-related risks, including many protection-related risks, since these are closely intertwined. It is worth noting that there are many other tools in the conflict-sensitivity toolbox beyond ‘Do No Harm’, and there is much practice and analysis that relates to conflict sensitivity more widely.

- The evaluation team could be perceived as gathering intelligence for one of the parties in conflict.
- Focus group discussions (FGDs) could become heated, or conversely reinforce divisions by missing opportunities to bring groups together in FGDs.
- The evaluation could become part of the battlefield for public opinion – people may respond in ways intended to promote a given agenda, raising concerns about both the credibility of findings and the ways that an evaluation report will be used in the future.
- The evaluation could present a biased analysis if it does not adequately present different stakeholder views.
- A predominance of views from more powerful/accessible informants may reinforce patterns of inequality and marginalisation.
- Contested conclusions or recommendations may contribute to increased tensions.

Undertaking an evaluation in a conflict-sensitive manner involves:

- assessing whether any steps in the evaluation process could contribute to tensions (this will need to focus on data gathering analysis and dissemination of the report in particular)
- carrying out new (or updating existing) conflict, context and stakeholder analyses, as this will inform the sampling frame and help identify possibly bias in the evaluation
- revising any planned steps in the evaluation in light of this analysis and situation updates to ensure they do not contribute to tensions (and where possible try to decrease them).

Safeguards and ethical considerations in data management

Several of the points above overlap with ethical considerations in human subject research more generally.⁴ These strongly apply when information on sensitive issues is sought directly from individuals who may have suffered harm or abuse, particularly in the area of sexual and gender-based violence (see 'Keep in mind' box on following page).

In addition to 'do no harm' considerations, **data-protection laws and confidentiality** must be understood and respected before entering into any data-management activity specific to evaluating protection in humanitarian action.⁴ To ensure the security and confidentiality of data (in use, in transit and at rest), there is a need to take technical and organisational security measures. Such measures may include the use of encryption software and the exercise of due caution regarding the use of remote-access databases. But it should also be noted that even encryption can be compromised if access controls are not well managed. Evaluators should consult the Protection Information Management system outlined in ICRC (2018), Chapter 6, and the Handbook on Data Protection in Humanitarian Action (Kuner and Marelli, 2017).

Confidentiality and its link to data protection are legal requirements and constitute an important part of the evaluator's ethical obligations. The protection of natural persons in relation to the processing of the personal data is considered a fundamental human right and freedom, as outlined in, among others:

- The Council of Europe Convention 108⁵
- Article 8(1) of the Charter of Fundamental Rights of the European Union
- Article 16(1) of the Treat on the Functioning of the European Union (TFEU)
- The Preamble to the EU General Data Protection Regulation.

The general principles underpinning data management and confidentiality in evaluation is that people 'own' their own life experiences and that attributable data is available to the evaluator only on a negotiated basis (Kushner, 2005: 74). Useful guidance on these points can be found in ICRC (2018), Annex to Chapter 6.



Keep in mind: Special considerations for primary data collection on sexual violence

The World Health Organization Ethical and safety recommendations for researching, documenting and monitoring sexual violence in emergencies (WHO, 2007) recommends that 'basic care and support for survivors must be available locally before commencing any activity that may involve individuals disclosing information about their experiences of sexual violence'. Evaluations are often assessing the availability of such services, and therefore these conditions can rarely be guaranteed in advance. Reflecting on the implications of this, Bain and Guimond (2014: 16) conclude that prevalence studies and other types of data collection conducted in the absence of GBV services should be seen as in violation of humanitarian ethics.

Any personal data or information used in the evaluation report should be used in a manner that complies with applicable laws on data protection, including respecting the principle of 'data minimisation', i.e. they are limited to what is necessary in relation to the purposes for which they are being used (e.g. 'An elderly widow from Damascus said...').⁶ A report may be sanitised and circulation limited to avoid spread of information that could be associated with identifiable informants, although some risk of leakage remains in any documentation.

There is also a need for extra care in data recording and data storage, as there is a risk of hacking, confiscation or data theft. Time should be invested upfront to establish the recording and storage system. A good practice from the handling of research involving HIV is relevant here – to pseudonymise cases and keep only random numbers on the files, storing names linked to numbers elsewhere. In a survey carried out in Nepal, all names were removed from all materials and kept in strictly confidential controlled files, while call records were kept in a separate place.

Extra security (encryption) measures may be needed for computers linked to the internet, particularly where data are stored and shared via remote log in.

Ensuring informed consent for primary data collection and use

It is important for evaluators to obtain meaningful informed consent from individuals regarding the collection and use of primary data.⁷ ICRC (2018) Section 6.9 provides a detailed explanation of informed consent in the context of protection activities and should be consulted by evaluation commissioners and teams.

Programme participants and other stakeholders contacted as part of an evaluation should freely consent to participate in the exercise without being pressured to do so. Informed consent also includes reassurance that declining to participate will not, for example, affect any services provided in such a case (Brikci and Green, 2007: 5). Obtaining informed consent before proceeding with gathering information requires evaluation teams to:

- ensure that all potential respondents, including children and young people, fully understand what is involved in their participation
- ensure that all participants are informed of their data-protection rights under the applicable law
- encourage questions and clarification before proceeding with interviews or other data-gathering exercises
- allow sufficient time for potential participants to reflect on and decide whether to take part and assess the respondents' understanding of consent by, for example, using quizzes, asking questions one-to-one or asking them to summarise what they have been told
- equip interviewers with information on services available to the interviewees (e.g. health and social services)
- be aware that some special considerations apply for all data-gathering exercises expected to cover issues or experiences relating to sexual violence (see 'Keep in mind' box opposite)
- let the participants know that they can withdraw from the data-collection exercise at any time.



Good practice example: Using the HESPER scale for informed consent before interviews

The HESPER Scale is a tool developed by the World Health Organization (WHO) with King's College London to look at perceived needs of people in crisis contexts across 26 'need items'. Ratings are made for each need item according to unmet need (or serious problem, as perceived by the respondent), no need (or no serious problem, as perceived by the respondent), or no answer (i.e. not known, not applicable, or answer declined). Respondents are also asked to name any other unmet needs not already listed.

The administration of the HESPER Scale by interviewers to respondents is preceded by an informed consent process. This is to ensure that respondents take part in the interview voluntarily, without coercion or fear that they will miss out on benefits if they do not participate, and to help avoid raising unrealistic expectations. Informed consent may be taken either verbally or in writing, depending on the context.

At a minimum, this should involve explaining to the respondent who the interviewer is and the agency he or she represents, the reasons for the survey, and an overview of the interview process, including the amount of time needed. Furthermore, it should be clarified that participation is anonymous, completely voluntary, that no compensation will be paid, and that there will be no benefits to respondents if they participate.

The interviewer should then answer any questions the respondent may have, before asking whether the respondent agrees to take part.

Ideally each respondent should be given a participant information sheet explaining all of the above (which they may either read themselves, or which may be read out to them), and each respondent should sign two copies of this sheet (one for the respondent to keep, one for the interviewer) as consent to take part in the survey. If the respondent does not agree to take part, he or she should not be pressured into doing so. Respondents should also have the right to withdraw from the interview at any point without having to give a reason.

Source: WHO and King's College London (2011: 24)

There may be cases where informants are suspicious of the evaluation team or data collectors, or where they may be comfortable only talking to outsiders rather than with people who are more involved in their lives. Key informants might not agree to sign any consent documents, believing that signing a document is a trick. The act of asking them to sign may stop them from engaging at all.

Pay attention to who collects and manages the data

In evaluating protection in humanitarian action, it is critical to pay attention not only to how data are collected and managed, but also by whom.

Recruiting local researchers (not just from the country but also from the locality) can enable access to data that would be too risky to collect. At the same time, however, this also reflects a transfer of risk where it is delegated to local partners and individuals. Ethnicity, gender, caste, religion and other factors can generate risks for both the interviewer and the interviewee.

All data collectors should receive substantial training in interviewing techniques, including knowing when to stop an interview (for instance, if someone else has entered the room or the informant is deeply distressed), being able to identify if approaching an informant would put the person at risk, and to determine an appropriate place to conduct an interview.

While expertise in evaluation and protection is essential, those with extensive experience in the geographical area are particularly crucial to evaluations in conflict contexts. They have the contacts and are more likely to obtain access, and to hear the authentic voices of the people on the ground. They can identify and navigate the stakeholders, contextualise informants and their biases, and distinguish exaggeration from fact. They can provide essential insights into the political economy of the institutional environment in which the intervention is being made.

Dealing with sensitive data: legal implications and political sensitivities

Data collected can have legal and disciplinary ramifications: Sensitive protection-related data, gathered as part of an evaluation exercise, can have legal and disciplinary implications. For example, there could be cases in which such data point towards criminal activities, violent acts liable to prosecution under national legislation, or malpractice and abuse (including sexual abuse) on the part of agency staff or partners. In addition, data-protection laws place additional restrictions on the use of special categories of personal data, such as data revealing personal identity. The obligations on the evaluator in responding to these situations will differ depending on the nature of the information, the jurisdiction in which alleged activities occurred, and the policies of the agency concerned. In some cases, there may be a clash of norms, whereby local law may force disclosure, or allow law-enforcement authorities to require the handover of data, while the agency's own policies defend confidentiality. In all cases, it is important that data-collection protocols clarify at the outset the evaluators' options and available channels when uncovering this type of information, including the conditions under which data should not be collected (e.g. if proper safeguarding cannot be guaranteed).

Data can touch on issues which are politically sensitive: A main challenge here is that evaluators may not be able to use all the information gathered during the fieldwork in their report. When drafting the evaluation report, this may result in some gaps in argument and evidence used to substantiate findings that were generated drawing from sensitive information. One report from Oxfam GB touches on this specific point:

Due to its sensitive nature, some of the material collected for this review has not been included in this report. The conclusions and recommendations aim to elicit important learning from the full range of experience including that which is not documented here, and the reader may find some disjuncture in making direct links from case studies to some conclusions and recommendations. (Oxfam GB, 2011: 37)

Evaluators should ask whether the programme or agency has established guidance on how to deal with sensitive data in terms of mandatory reporting and disclosure of sensitive data in the final evaluation report.

9 Data sources and constraints on data management

Data-gathering and management efforts should be informed by applicable data-protection laws, and an awareness of a range of risks and constraints, some of which can be planned for and mitigated. At the very least, potential scenarios should be considered when planning data management and a 'do no harm' approach should be applied. The following factors should be considered in making decisions about sources and constraints.

Insecurity and constrained access: Insecurity means that evaluators tend to make short visits to places that are easier to reach. Sometimes interviews cannot be pre-arranged as this would create risks, so the evaluator can only speak with anyone who happens to be there at that time (a 'convenience' sample). Potential informants often distrust outsiders and are reluctant to talk. This makes it harder to draw out data, and also to achieve and document 'informed consent'. Interviews at more secure places where people gather (e.g. markets) may somewhat reduce these risks.

Access by international members of the evaluation team may be severely restricted: This can in turn lead to reliance on more junior evaluation team members with little experience or training.

Trauma and shame: Asking people to describe traumatic experiences can re-traumatise them. They may feel shame about the experience, particularly if they have experienced sexual violence. As noted below, alternatives to collecting data from traumatised individuals should be considered.

Creating or aggravating risks for informants: Informants may have a well-founded fear of reprisal for disclosing information about their experience. This relates not only to what is written in the report, but also to secure storage of data. The conclusions or recommendations made in the evaluation report, if made public, could inadvertently contribute to tensions and thus increase vulnerability. Even if transparency is a general principle to strive for in evaluation, the special circumstances surrounding protection may suggest that some reports must remain confidential.⁸

Poor data environment: Even where there are baseline data and indicators, a rapidly changing environment (including rapidly fluctuating populations due to displacement and/or cross-border movements) may mean that baseline data and indicators are no longer relevant, or that indicators have not been regularly monitored. Official records, including national statistical data or secondary sources, are often weak or non-existent, which will affect the choice of sampling frame. The political sensitivity of key variables may have prevented data from being collected or negatively affected its credibility. There may also be challenges in determining what makes a 'typical' case for case study selection.

Polarisation: In situations of conflict, views tend to become polarised, which in turn heightens the risk of bias. The evaluation itself can become part of the battlefield for public opinion as informants or stakeholders respond and act strategically – trying to use the evaluation to support a particular policy narrative regarding the causes of conflict or donor responses to it (sometimes referred to as 'policy-based evidence formation'). Similarly, there can be issues with bias and polarised views regarding the evaluation itself. For example, implementers, donors and the evaluation community may hold differing and contested views on the feasibility, ethics and appropriateness of using a given standard for evaluation designs.

More limited use of common data-collection tools: Certain data-gathering tools, such as FGDs, may have more limited application in evaluations that look at protection in humanitarian action because of issues relating to stigma and fear of recrimination.

Table 2: Data types and sources, their relative strengths and weaknesses and vulnerability to bias

	Examples	Possible use in evaluating protection in humanitarian action	Possible constraints and vulnerability to bias
Primary data generated from individuals	<ul style="list-style-type: none"> • Incident reports • Eye witnesses (including through mobile phones) • Testimonies • Surveys 	<ul style="list-style-type: none"> • Magnitude of violence • Types of incidents • Perceptions of safety and security 	<ul style="list-style-type: none"> • Individuals may fear sharing information, even to those deemed 'local', due to fear reprisals or stigma • Quality of data is highly dependent on skills of the interviewers / researcher / or evaluators gathering the information
Secondary data from local stakeholder groups, CSOs, NGOs and other international actors	<ul style="list-style-type: none"> • Focus group discussions • Panels • Surveys • Monitoring reports • Self-reporting / self-assessments reports 	<ul style="list-style-type: none"> • Perceptions of safety and security • Understanding and contextualising perspectives, attitudes and behaviour in the affected populations and programme participants and how these change over time • Reconstructing / validating / testing logic models and theories of change • Understanding anomalies / outliers in survey results • Illuminating cases selected in purposeful sampling approaches 	<ul style="list-style-type: none"> • Social desirability bias • Group effect bias that may skew results towards uncontroversial and commonly held views • Conscious partiality of data providers
Secondary data from official and administrative sources	<ul style="list-style-type: none"> • Police records • Court records • Hospital records • Morgue records • Demographic and health surveys 	<ul style="list-style-type: none"> • Number of crimes • Deaths • Violent events • Prevalence studies • Incidence rates of reported domestic and sexual violence 	<ul style="list-style-type: none"> • Often unavailable – inaccessible, infrequent, inconsistent, lack internal validity (i.e. is the same thing being measured over time using the same set of measures) • Lack of reporting (due to stigma, recriminations, or discretionary use of power by law-enforcement officers) • Bias from usage: they only capture cases that have been reported or detected or for which services were sought* • Poor state capacity to collect data • If the official data set being used is of questionable reliability then caution is needed in how it is used – the evaluation report could be quoted out of context, and give extra credibility to an unreliable source data set
Secondary data sets from occasional country-specific data sets	<ul style="list-style-type: none"> • Periodic country-specific perception surveys • Ongoing periodic country-specific surveys • E.g. Small Arms Survey multi-year 'Sudan Human Security Baseline Assessment. • Event monitoring mechanisms • E.g. Risk Management Office established by DFID and GTZ in Nepal (now discontinued) collected data from programme monitoring reports, staff reports and media monitoring 	<ul style="list-style-type: none"> • Type of incidents by geographic location in a country • Trends in violence episodes • Perception of safety • Dataset on violent events 	<ul style="list-style-type: none"> • Sporadic release and update • Data gaps • Lack internal validity (i.e. is the same thing being measured over time using the same set of measures)

Sources: Compiled and adapted drawing from Hext Consulting (2012) and Church and Rogers (2006: 206-210)

* **NOTE on administrative and official records of incidence rates:** they are particularly vulnerable to usage bias because the data may show an increase in violence although actual incidence rates could be decreasing. This could result from an improvement in information systems, from improving levels of trust in the police or other reporting systems. For these reasons, the data may show increasing levels of violence that previously went unreported.

Insufficient access to a representative sample: Factors such as limited time, logistical or security constraints, or even uncertainty about who is affected by protection concerns in the overall population, may limit the extent to which an evaluation team can plan for and collect data that are sufficiently representative to draw generalisable conclusions about target populations. These risks can be mitigated by careful planning to know where people are likely to be and when, and to take into consideration issues of gender and ethnicity that can compromise access to different populations when deciding on the composition of the evaluation team.

Selection of the different primary and secondary sources of data in evaluating protection in humanitarian action should consider their possible advantages and disadvantages and their vulnerability to different types of bias. In addition, as in all aspects of evaluation, it is always important to be transparent in identifying and acknowledging gaps and biases in data collection.



Good practice example: Ethical procedures

Ethics standards for research on human subjects require that a caregiver give some form of permission for a child to participate in research, with exceptions made only in extreme circumstances.

In a pilot study in Rwanda to develop a child protection index to measure the strength of a child protection system, data collectors were trained to be aware of the effects that questions may have on the respondent and how best to respond, based on the respondent's level of distress. They were instructed not to provide counselling, but to inform respondents of services available and how to obtain access to them if necessary.

The Association of Volunteers in International Service-Rwanda and UNHCR Rwanda agreed to exempt researchers and data collectors from any existing mandatory reporting policies on abuse and violence. When a case was identified, the respondent was informed of services, and asked if s/he would like assistance in obtaining access to them.

Source: Meier, Muhorakeye and Stark (2015: 35-37))

Special considerations in primary data collection

Primary data collection about protection incidents and overall incident rates is particularly contentious. Evaluators should proceed with great caution before deciding to collect primary data about protection incidents and consider asking why such data are needed in the evaluation. Further, incidence rates can be seen as more pertinent to programme design than to evaluation, so it may be inappropriate or unnecessary to collect such data for evaluation purposes if the programme has already done so.

Where there is a high risk of re-traumatisation or reprisals, there may be serious repercussions for informants. Ethically, it may be indefensible to ask an individual about episodes of violence and trauma if the evaluator does not then link the person to a counselling service (see ICRC, 2018, Section 5.5 for more information). Since it is generally beyond the scope and capacity of the evaluation to establish the necessary protocols and safeguards, it is essential to engage with the organisation being evaluated to address this risk.

There are some alternatives to collecting primary data from individuals. For example, incidence rates could come from other sources – such as service providers (as illustrated in 'Keep in mind' box opposite) or from those engaged in advocacy on the same issue. Below are some suggestions for the measures that can be considered when data collection touches on protection incidents:

- engage those already providing a relevant service to conduct the data collection, as they are able to link the informants to the service
- rather than asking people to recount their own experiences, ask about someone else who has had this sort of experience (mother, sister, etc.)
- consider using interviewees with some basic counselling skills
- at a minimum, ensure that interviewers have sound interview skills.



Keep in mind: Challenges and constraints in gathering and using GBV data

'GBV is difficult to quantify as many cases go unreported, its scope is difficult to estimate and existing data is often misunderstood, misrepresented and ineffectively utilised. ... Prevalence studies can provide some idea of the overall picture of GBV in a country or area. However, they are only estimates and generally provide little information more subtle or short-term changes in GBV trends.'

Source: Bain and Guimond (2014: 16-17)



Good practice example: Conducting surveys in access- and data-constrained environments in conflict and post-conflict settings

A survey was conducted in north-east Afghanistan in a situation where official population data were not available. In order to develop the sampling frame, interviews were held with the village council (*Shura*) to determine the number of households in the village, which was used to calculate the number of interviews to conduct in that village. Some areas had no maps at all, and not even agreed names for villages, making it very time consuming to determine a sampling strategy.

The same survey commissioned quarterly reports on communities and districts in which the survey was conducted to capture significant local events significant changes, disasters etc. which helped to identify contextual factors which were key in analysing the survey findings.



Random sampling emerged as the best practice in north-east Afghanistan:

We opted to collect a random sample of households in every community, for every survey. An alternative would have been to collect panel data – that is to sample the same households for both surveys. We did not do this because we were afraid of high attrition. We anticipated that a deteriorating security situation would have forced many households to flee or be on the move for work. We also wanted to minimise risks for our respondents, households who speak too often to foreigners might have been at higher risk of reprisal by insurgents, which could in turn affect our responses. (Böhnke, Koehler and Zürcher, 2014: 112)

The survey team found these measures enabled a strong response rate, and thus a strong survey quality:

Our results are also consistent with the conclusion that different dimensions of armed conflict can affect survey data quality in opposing direction, with higher numbers of bombings in the local area slightly increasing refusal rates but higher number of nationwide political events actually increasing response rates through decreasing both refusals and non-contacts. (Axinn, Ghimire, and Williams, 2011: 23)

Consider alternatives to primary data collection from individuals

For the reasons discussed above, the need to look for alternatives to collecting primary data from individuals is particularly high in evaluating protection in humanitarian action. Data about incidents can also be extrapolated through exploring proxy indicators.

Data sources beyond interviews with individuals are often overlooked. However, there is some untapped potential for drawing on official data, which is a growing area in monitoring and evaluating conflict-related crime and violence programming (CCVR, 2012).

Some humanitarian responses occur in countries with reasonable state capacity, such as Indonesia, Lebanon and Philippines, where official data may already be of good quality. In other cases, there may be ongoing donor investments to improve state capacity in establishing official data in areas of relevance to protection.

Consider gathering and using primary data from service providers

Gathering and using primary data from service providers is seen as good practice in protection-related programming as well as in evaluation.

For example, a service provider that classifies stages of post-trauma healing could report on how its users are progressing through the different stages. However, service providers sometimes lack capacity for data collection, and collecting data for specific evaluation purposes may be a low priority. Ideally, the intervention being evaluated may have capacity-development components to enhance service providers' monitoring capacities, but this is seldom the case in humanitarian interventions.



Good practice example: Gathering M&E data through local partners

In DRC, Search for Common Ground partnered with and supported local women's groups to provide services for women, and also trained them in gathering data for M&E purposes. That appeared to have enabled women to talk more openly about taboo subject matters because it was a case of local women talking to local women and being then linked to locally available services.

Evaluators (and especially those responsible for strengthening monitoring systems) may therefore need to consider including some elements of capacity strengthening whenever service providers are expected to take part in data-collection work that is specifically commissioned for an evaluation. There is also a risk of bias, particularly if the service provider is directly supported as part of the programme being evaluated. There may be also options in terms of peer-to-peer data collection, as described in the following example from Search for Common Ground.

Evaluators may invest in an analysis of the context and the interests of different stakeholders to help mitigate this bias. Even literature reviews – as suggested in the evaluators' insight below – can be helpful in this respect (see [Toolkit item # 5](#)).



Evaluator's insight: Dealing with bias

In order to cope with the expected bias from informants, the multi-donor evaluation of conflict prevention and peacebuilding programming in South Sudan used the field-based interviews to verify the literature review rather than using them as the key data source:

Importantly, the evaluation was not dependent on these field level interviews and discussion groups - which might be seen as partial or biased – but rather these were used to triangulate the more substantial evidence and preliminary findings from the first stage literature review and analysis. (Barnett and Bennett, 2014: 45)

10 Data on the less tangible and harder-to-measure dimensions of protection in humanitarian action

Data on perceptions

Many in the conflict-prevention and peacebuilding fields are very confident in using perception-based data – and many would claim that ‘perception is as important as reality’. Individuals are often motivated to commit violence on the basis of a perception – how they view another group, or rumours about tragedies – and indeed much peacebuilding work is about trying to change such perceptions. Some evaluation users, however, may be sceptical and see data regarding perceptions to be a weak proxy for ‘hard data’ on actual violence, displacement or other variables.

The security and justice field also works with perceptions – particularly perceptions of safety and security. It can be more politically and technically viable to collect perception measures of safety and security (such as the percentage of men and women who fear a crime) than incidence data (number of incidents of violent crime). Examples of data-collection tools focused on perceptions that could be useful here include:

- movement maps that visually capture where people feel safe to move, possibly mapped over time to show changes in perceptions of safety and security
- body images, where women are invited to talk about GBV through drawing bodies and describing what they are most and least proud of in their own bodies.

It should be noted that such tools are highly specialised and using them effectively and sensitively requires a significant investment in training and engagement of appropriate team members.

Perception-based data should not be used in isolation.
Three considerations apply here:

- Data should be triangulated and analysed together with other sources of evidence about the programme and/or the context (for example, media reports of violent events can be a good corroborating indicator).
- The utility of perception-based data depends on the degree to which changed perceptions (for example, of gender roles and violence) were part of the programme objectives.
- Perceptions and incidence data may not align: for instance, there could be a time lag between an actual improvement in crime statistics and perceptions reflecting this. There may also be significant variations across short geographies: a village that suffered an atrocity will have a very different sense of security that one nearby that did not.

As in other fields of humanitarian action, many evaluation stakeholders have less confidence in measuring perceptions. Actors accustomed to relying on 'hard data' on malnutrition, litres of water available per person or disease vectors may be inherently sceptical of perception data.



Keep in mind: A sample of perception-based indicators

An Itad report on behalf of DFID, assessing a suggested list of governance and conflict indicators endorses several perception-based indicators (for use in conjunction with objective indicators), some of which are relevant to protection:

- percentage of citizens who say they feel safe going out in their neighbourhood at night (disaggregated)
- percentage of citizens who believe bribes are necessary to access police services
- percentage of target population who report positive attitudes to civil–military relationships and to reintegrated combatants
- percentage of community members who do/don't feel threatened by the presence of ex-combatants.

Source: Barnett, Barr, Duff and Hext (2011)



Good practice example: Measuring community security – Saferworld

Saferworld implements a large spread of community security programming globally. It promotes a participatory approach to monitoring and evaluation, with the communities themselves determining the dimensions to be measured. Saferworld's Community Security Approach addresses insecurity at three levels and conducts measurement at all of them: community/local level, sub-national/district level, and national level. The measurement of community security combines both perception-based data (e.g. sense of security) with more tangible dimensions of security (e.g. number of attacks on community). These measures span all three levels.

Community / local level

As part of the programme, the community identify security concerns and generate action plans, and determine how to monitor progress and measure success. Key dimensions to change that should be measured at this level include:

- specific outputs to be achieved (e.g. establishment of local police post)
- changes in the way the community feels about itself and agency (e.g. willingness to tackle sensitive areas of concern)
- changes in the perception / sense of the community about their situation (e.g. do they feel safer? Do they trust their authorities?)
- changes in relationships within the community and/or with others (e.g. relationships with the police service)
- changes in the behaviour of the communities and the security provider.

Sub-national and national level

The programme uses research and advocacy to link local improvements up to sub-national and national levels to promote policy change that draws on the local programme experience. Key dimensions of change that could be measured at this level include:

- behaviour of security providers towards communities (e.g. number of attacks by security providers on individuals and/or communities, the extent to which security providers see themselves as a service to the community, rather than a force for control)



- community behaviour towards security providers (e.g. willingness to report crime or security issues to relevant authorities)
- relationships between the community and security providers (e.g. quality of interaction between security providers and communities in meetings, level of continued reliance by communities on non-state, informal security providers)
- feelings of safety and security (e.g. the proportion of women who feel confident of walking in the community after dark)
- changes in the way sub-national and national security providers consult, engage and respond to communities
- changes in how security budgets are defined and used.
- the programme uses a participatory evaluation process in which the community convenes for a day to identify transformations in the relationships and behaviours behind insecurity, and how these changes have affected peoples' experience of security. Annual community security assessments have been a valuable data collection instrument for the programme which identify the nature of security in that locality and can track specific security issues, the availability of services, and the feelings of safety of the communities.

Source: Saferworld (2014)

Sense of security

The sense of security is a particularly important aspect of judging outcomes related to the environment for protection and the perceived relevance of interventions and may even help understand the sustainability of the changes it induced.

Security indicators need to be developed early in the programme so they can be used in monitoring. Without a baseline and monitoring data, it is likely to be too late to make meaningful measurements in a summative evaluation. Indicators are best developed as 'participant-generated indicators' by asking people in the community about what they believe indicates that the situation is safe. For example, in Darfur IDPs stated that they would send one or two family members back to their place of origin and wait to see if they could remain safely throughout one agricultural season.

All such participant-generated indicators should be disaggregated by age, sex, and any other salient distinctions.

A community security assessment or FGDs in same-sex groups can also be used. A basic question would be 'under what conditions would you feel safe doing xxx activity' (e.g. collecting firewood or earning an income or moving between location A and B etc.).

Such data can contribute to a formative evaluation, providing insight into the relevance of the intervention.

Endnotes

- 1 Available at: <https://www.icrc.org/en/handbook-data-protection-humanitarian-action>
- 2 These points are also covered in the United Nations Evaluation Group (UNEG) guidance on ethical obligations to those to initiate, manage and carry out evaluations (UNEG, 2008). These obligations include: respect for dignity and diversity; human rights; confidentiality; and avoidance of harm.
- 3 This is a core area of evaluative work in the realm of peace-building evaluation, evaluation of conflict transformation and evaluation in the security and justice sectors.
- 4 In the context of evaluation, ethics has been defined as encompassing concerns about the rights, responsibilities, and behaviours of evaluators and evaluation stakeholders (Yarbrough et al., 2011: 106).
- 5 In many respects, the Child Protection community has paved the way in researching and clarifying informed consent procedures when working with children, including in emergencies and crisis settings. For example, UNICEF released an online resource portal called Ethical Research Involving Children (ERIC) (<http://childethics.com>) to compile and make accessible the latest resources and expert thinking about key ethical issues involving children and how these might be addressed in different research (and evaluation) contexts.
- 6 The Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, available at: <https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108>
- 7 See ICRC (2018), Section 6.11. for more information on data minimisation for protection activities.

- 8 It is important to identify the most appropriate legitimate basis for collecting and processing personal data. Consent is only one of the possible legitimate bases for collecting data but will almost always be the most appropriate basis of legitimacy in evaluation contexts. For consent to be valid, it must be informed and freely given. This means that the consenting party must be able to understand what they are consenting to and how the information they provide is going to be used. Outside the evaluation context, there may be instances where personal data is collected on a different basis of legitimacy. For example, in the collection of data from unaccompanied children, who may be unable to understand the future use of their data and who do not have a legal guardian who can consent on their behalf, it would not be appropriate to rely on consent as the legitimate basis. In such instances, data collectors must identify a legitimate basis for data collection other than consent, such as vital interests, public interest, legitimate interests, compliance with a contract or with a legal obligation. Without a legitimate basis for doing so, personal data must not be collected. Although the norms and standards that guide data collection for evaluation are, arguably, different from those that govern the conduct of protection actions, evaluators must also reflect upon the legitimate bases for data collection in compliance with applicable personal data protection laws and policies. ICRC/Brussels Privacy Hub (2017) Handbook on Data Protection in Humanitarian Action provides useful guidance on this issue, and evaluators should refer to this before conducting evaluations.

Notes

C / Analysis



C / Analysis

Content of this module at a glance

Section	What this section does	Who this section is for
11	suggests how evaluation teams should revisit the original intervention logic as a point of departure for their analyses	primarily evaluation teams
12	reviews the concepts of causality, attribution and contribution and how they are likely to be applied in EHA protection	primarily evaluation teams, but also useful for evaluation commissioners when assessing the quality of evaluation reports
13	presents insights for evaluating advocacy and other initiatives intended to influence the protection environment	primarily evaluation teams

11 The starting point for analysis: Revisiting the intervention logic

In evaluating protection in humanitarian action, challenges often arise when reflecting on the intervention logic or theory of change (ToC) (see [Toolkit item #4](#)). Uncertainties about the intervention's approaches need to be unpacked and reconsidered.

We start this module with a discussion of a number of fundamental factors that may skew analyses and of which evaluators should be aware:

- **Weak designs and under-developed intervention logic:** Projects and programmes developed hastily in response to an acute crisis may lack a ToC or other intervention logic that articulates how activities and outputs were expected to address protection risks, much less the assumptions about how the initiative was expected to overcome obstacles inherent to the conflict context. An evaluator is sometimes tasked with providing guidance for a revised or enhanced ToC for future programming.
- **Influence of institutional worldviews:** Analysis of protection risks and needs in a given context may be driven more by institutional worldviews or prevailing policy narratives than by deep contextual knowledge, thus skewing the assumptions underpinning programming. An agency may ignore potential mismatches between programming and protection risks if it has operated under a 'if all you have is a hammer, everything looks like a nail' approach to programme design: providing the goods and services they normally supply, rather than those required in the situation. This may sometimes limit the parameters of the evaluation when ToR fail to provide room for questioning the unconscious worldviews that frame programming. This can even lead to a narrow evaluative focus on whether the intervention is 'doing things right' (within standard agency modalities) when there are major unresolved questions regarding whether the intervention is 'doing the right thing' (in terms of protection risks). For this reason, an evaluation with a narrow focus on effectiveness may draw an entirely different conclusion to one focused on broader relevance.
- **Complexity and a focus on 'doing':** If there are multiple components or different protection actions nested within a larger intervention, there may be a lack of detail on what is actually expected to be implemented for whom, by when, where and how. As a result, there may be undue influences or deep-seated bias affecting what is implemented, assessed, measured and monitored.

- **Perceptions and more hidden dimensions of results:** A sense of safety or security is much harder to describe and measure as it has more 'hidden' components than other, arguably more tangible areas of results linked to assistance provided in terms of health, nutrition or sanitation.

For these reasons, among others, evaluators should be aware of some key questions:

- Be alert to the different ways in which agencies describe and frame protection. Note that some agencies may use similar terminology to mean different things.
- Be aware that there may be some disconnect in how different teams and programmes within the same organisation – and even within the same operation – see themselves working within and around protection, and how this may influence the logic of the intervention. This is particularly important in interventions where protective actions are being implemented by staff from other sectors.
- Be aware that different ways of approaching protection in humanitarian action are likely to co-exist in the same programme, intervention or context. This is likely to complicate analysis because it affects the extent to which evaluators will be able to identify and 'isolate' the specific elements/factors contributing to protection outcomes of interest.
- Be alert to how the international legal frameworks applicable to the different contexts and situations in which humanitarian actors operate (e.g. in international and non-international armed conflicts and other situations of violence) can affect the relevance, feasibility, connectedness, coherence and effectiveness of certain types of protection actions.

By unpacking the often diverse implicit and explicit expectations and assumptions across the results framework, analysis of intervention logics and theories of change can also help focus lessons and recommendations in ways that resonate with different users' decisions about whether, where and how the intervention could be scaled up or carried out in other settings under specific conditions (Bamberger, Rugh and Mabry, 2012: 183, 227).

Even if there are dangers that a focus on particular ‘results’ narrows the perspectives of certain users, analyses of such results can be used as entry points for a broader discussion about the protection-related issues that impinge on the targeted actions.

Specific uses of programme theory in evaluation that can boost the explanatory strength¹ of an evaluation include drawing conclusions regarding:

- the strength and weakness of the intervention logic underlying the design of the intervention
- the strength and weaknesses of how the intervention was implemented
- how contextual factors contributed to, or militated against, the achievement of intended results, thereby drawing attention to the contextual relevance of the intervention modalities
- how the intervention affects, and is affected by, different groups, for example the extent to which ‘do no harm’ principles were applied and due attention given to gender perspectives and issues related to marginalisation and vulnerability (Bamberger, Rugh, Mabry, 2012: 182).

Critically reflecting back on the programme theory can also be important when considering whether there may be alternative explanations for the changes in, for example, the perceptions of safety in a community or the actions of authorities.

12 Analysing causality, attribution and contribution

Causality

The ALNAP EHA Guide presents an overview of critical issues for evaluators when analysing evidence for causal questions. Section 16.4 presents eight strategies for establishing causal inference that can be applied to all types of evaluation. Evaluation teams are encouraged to review this text in addition to the more protection-specific guidance that follows.

Analyses of causality need to start with describing the baseline, which could include existing levels of service provision, processes already underway, the current situation of human rights abuses, the institutional setting, amongst others. Ideally this will have been done as part of planning the intervention, but in humanitarian settings this step is often forgotten or undertaken in a rudimentary manner, which can create an additional burden on evaluation teams who must then ‘reconstruct’ the baseline by drawing on a range of secondary data or stakeholder recollections.

Analyses also need a description of the protective actions being taken, including inputs, activities and outputs. This may also include description of internal constraints (e.g. budget, human resources) and external factors and events impinging on the intervention.

The next step is to assess the relational assumptions in the programme’s explicit or implicit intervention logic. This involves establishing whether a relationship between two or more phenomena is assumed to exist and, if so, its direction and magnitude. The empirical data gathered in the evaluation itself may significantly change the evaluation team’s understanding of who holds these assumptions and how they are interpreted in practice.

Most evaluations are steered by normative analyses – that is, they compare ‘what is’ with ‘what should be’ and the current situation with a specific target, goal or benchmark. However, many humanitarian interventions are focused heavily on delivering a set of outputs, with grander normative objectives described in somewhat vague or visionary terms. This means that the evaluation team may also need to reconstruct a more realistic ToC based on either (a) the intervention logic as perceived by key stakeholders in the programme, or (b) the evaluation team’s or commissioning officers’ own assessment of what would constitute a more plausible ToC.

Analysing quality and value – evaluations almost invariably ask about the overall conclusion regarding whether an intervention can be considered a success, an improvement compared to the previous situation, or the best option (Rogers, 2014: 10). Some authors refer to this as asking truly evaluative questions (Davidson, 2014: xii) to underscore that what makes evaluation different from other endeavours (e.g. performance measurement and monitoring) is asking ‘how good’ and ‘how valuable’ the results of a certain intervention are for specific groups and individuals and why. In this sense, truly evaluative questions are those that do not stop at asking ‘how things have changed’ and ‘to what extent has the change been brought about the intervention being evaluated’ but also examine the importance, quality and value dimensions of change.

Causation (or causal) analyses seek to establish the intervention’s role in producing the results described or implied in the (reconstructed) intervention logic. One central concern when answering causal questions is documenting that a given result, change or effect has been caused by the intervention and not by coincidence or by other concurrent factors. Particularly in the complex and dynamic contexts that characterise humanitarian emergencies, it is essential that the evaluation does not equate correlation (e.g. a change in the frequency of protection violations) with causation. Contribution analysis (discussed below) is a way to unpack the question of relations between interventions and actual phenomena.

Special considerations apply when analysing causality in programmes and interventions in which protection is more implicit or has been woven into other (non-protection-oriented) services, activities and programmes. Establishing causality is likely to be a more complex and resource-heavy exercise because of the work needed to identify and reconstruct the ‘bundle’ within which protective elements have been included.

Attribution and contribution

Establishing causality is not straightforward, particularly in crisis, fragile, conflict and post-conflict settings, and there are different ways of looking at causation (see [Change-centred approaches in Toolkit item #3](#)). One of the pervasive challenges to establishing causation in evaluation is that it may not be possible to isolate the results brought about by a given intervention among a host of other contextual factors. This point is commonly referred to as the attribution problem. This point is even more pressing in the case of protection where, as noted by the ICRC (2018, Chapter 2), ‘it is not likely that a single type of activity can achieve comprehensively reduced risk. Even within one organisation, achieving reduced risk may require a variety of sectors and disciplines working towards a common desired outcome’.

Attribution requires establishing the causal implications of an intervention and/or the causation of an observed phenomenon (Scriven, 2010: 1; also see Gerring, 2012). However, especially in the context of evaluating humanitarian action, it is rare for causal attribution to refer to sole attribution. Rather, it often refers to establishing partial attribution or analysing contribution to impacts² (see [Toolkit item #6](#)).

There are various techniques to help evaluators examine causality – whether in terms of sole attribution, partial attribution, or contribution to results (drawing on Chigas, Church and Corlazzoli, 2014: 20). At the broadest level, analysis and techniques used to answer causal questions in evaluation will pursue one or more of the following:

Factual analysis involves asking: what kind of results and changes (outcomes or impacts) occurred for whom in a given context? How did the actual results of the programme or intervention compare to those expected from the logic model or ToC that informed the intervention? Are the results and changes that can be observed consistent with the theory?

Analysis of alternative explanations involves examining different scenarios posing alternative explanations (other than those related to the outputs of the intervention) that could account for the observed changes and results.

Counterfactual analysis produces some estimates or seeks to explain what would have happened if the intervention had not occurred. Conventional attribution analysis requires the group receiving the programme or intervention to be matched to a comparison group. Here, there are stringent requirements for dealing with bias³ and for dealing with alternative explanations of the observed changes, which ideally should be eliminated by using a counterfactual analysis. While common in many forms of evaluation, rigorous counterfactual analysis is rare in evaluation of humanitarian action.

Contribution analysis seeks to assess the extent of the influence of a particular actor in contributing to the overall changes resulting from a collaborative intervention (Bamberger, Rugh, Mabry, 2012: 404). Contribution analysis (Mayne, 2001) also refers to a specific technique used to establish contribution in a structured manner, following six steps:

1. develop the theory of change
2. assess the existing evidence on results
3. assess alternative explanations
4. assemble the performance story
5. seek additional evidence
6. revise and strengthen the performance story

13 Insights from evaluating advocacy and other initiatives intended to influence the protection environment

Evaluating advocacy⁴ in aid and development settings is a growing area within the broader practice of evaluating advocacy, policy influence, communication and campaigning (for an overview see, for instance, Stachowiak (2013), Tsui, Hearn and Young (2014) and LFA (2013)). With a few exceptions, there is a dearth of specific guidance for humanitarian evaluation looking at advocacy. Many of the challenges associated with evaluating advocacy efforts can be found in other contexts, but several are more prominent in humanitarian contexts.

UNICEF has made a sizeable attempt to document, systematise and produce guidance on monitoring and evaluating advocacy, including a specific section on advocacy in the context of crisis and emergency and post-emergency contexts (Coffman, 2010). Its guidance on monitoring and evaluation of advocacy describes those challenges as:

- **The speed of decision-making and the urgency of information needs:** During and after an emergency, a quick systematic assessment is necessary to inform decisions regarding advocacy efforts, but the nature of such settings often impedes this.
- **Inherent volatility and complexity:** Due to the volatility and complexity of emergency and post-emergency settings, it can be hard to identify advocacy targets. 'This poses difficulties not only in conducting advocacy in the first instance – and hence in demonstrating its effects in light of a rapidly changing landscape – but also in accessing the most qualified stakeholders who can shed light to the evaluation team on UNICEF's efforts' (Coffman, 2010: 14).
- **The abstract nature of advocacy processes can make data collection difficult:** Advocacy processes have abstract outcomes that are difficult to define precisely (public will or political will, for example). As such, less conventional methods are applicable to evaluating advocacy efforts (Coffman, 2010: 20).



Keep in mind: Comparable challenges in evaluating advocacy and influencing work

Evaluators of protection in humanitarian action could benefit from looking at the practice of evaluating advocacy and policy influence in the broader aid and development settings. The two domains grapple with a comparable set of evaluation and measurement challenges, including:

Causal relationships: Linking advocacy and outcomes is complex.

Subjective gains: Defining success is challenging and varies depending on who is asked. The goal posts can often shift depending on the circumstances.

Multiple approaches: Influencing policies and behaviour change can be part of many approaches including lobbying, advocacy, policy research or campaigning. It may be difficult to assess which approach leads to which results at the outcome and impact level.

Programme approaches are inherently more speculative than direct interventions, and the benefits are less easily articulated, typically take longer to achieve, and are also less easily assessed or measured.

Long horizons: Advocacy and influencing work are long term. Change can be slow and incremental.

Conflicting political process: Influencing often means engaging in a process that may have political consequences, which in crisis and conflict situations may be even more far-reaching and draw the evaluation into sensitive and contested areas related to humanitarian neutrality and impartiality.

Tension about metrics: There is a tension between the desire for 'metrics' or quantifiable indicators and the need for usefulness analysis of progress. Many metrics are either too narrow or short term, focusing on activities such as the number of newspaper citations, or too broad or distant, such as changes in policy or legislation.



Focus on measurable data in advocacy evaluation (e.g. data from social media, news stories) tends to be far away from the real value and actual change, and so is comparatively uninteresting for users when there are demands for evidence that advocacy and campaigning is achieving tangible results.

Most outcomes and impacts are hard to see: In value terms, advocacy and campaigning is like an iceberg: most of the impact may be submerged and hard to see. The temptation to focus only on the part that is visible risks creating a radically false picture that generates misleading information and so encourages poor decision-making.

Sources: Coe and Majot (2013); Reisman et al. (2007); Chapman and Wameyo (2001); Tsui, Hearn, and Young (2014); Schlangen and Coe (2014)

Above is a compilation of insights into evaluating advocacy and other initiatives intended to influence policy. The intention is to highlight some of the emerging learning and guidance in that field that may resonate with those evaluating protection in humanitarian action.



Evaluator's insight: Move towards a 'try and evolve' approach to monitoring and evaluation

Snowden and Boone (2007) have suggested that the appropriate management style for complex interventions is to use an experimental 'try and evolve' approach, which recognises that even successful interventions will involve missteps or mini-failures. In that light, identifying and learning from these missteps is essential to guide programming and should not be understood as lack of effective planning and design – as it is the case in some 'conventional' monitoring and evaluation guidance (Tsui, Hearn and Young, 2014)



Evaluator's insight: Focus on evaluating progress and contribution

Advocacy strategies evolve, which means that activities and desired outcomes also change. Course-correction and adjustments are the most realistic expectation in monitoring and evaluating advocacy. That is one of the reasons why evaluating progress is also important. 'Advocacy M&E typically focuses on the advocacy journey rather than just the destination' (Coffman, 2010: 2). That journey has usually started before the intervention and will continue long afterwards. Evaluations should recognise this broader perspective at the outset. Drury (2014: 27), for instance, offers a clear example of contribution analysis applied to protection advocacy in a humanitarian setting.

In advocacy evaluation, there is a strong focus on articulating and measuring interim outcomes because ultimate goals (e.g. passing a resolution or changing an entire policy approach) can have very long time horizons – years or even decades. An important focus of advocacy evaluation, therefore, is interim outcomes, which (LFA, 2013: 5):

- Are benchmarks or milestones that demonstrate incremental progress towards your ultimate goal (e.g. getting an important policy-maker on board as a champion).
- Can be the direct outcomes of your advocacy activities or tactics (e.g. after meeting with important policy-makers, they commit to introducing a parliamentary bill).
- Are often outcomes that you must achieve in order to reach your ultimate goal (e.g. you need a certain set of policy-makers on board in order to get a bill or a resolution passed).



Evaluator's insight: Distil evaluative information meaningfully

Distilling information down to the basics is an appealing, efficient and necessary way to communicate what happened – particularly if it can be done with numbers. The risk is that numbers, rather than being an aid to strategic decision-making, might substitute for it. As a rule of thumb, the more complex the context being assessed, the less credible meaning is to be found in its simplified distillation. So, for organisations trying to assess the value of advocacy and campaigning, translating qualitative information into numbers can devalue this information by stripping it of the very detail that gives it value. It also typically conveys a false sense of precision and objectivity. For this reason, one working principle in reporting advocacy and campaigning should be 'no narrative-free data' (Schlangen and Coe, 2014: 7), a principle that is highly relevant for evaluating protection in humanitarian action.



Evaluator's insight: Use single and multiple case studies

Case studies often examine different aspects of an advocacy effort and collect data from a wide range of stakeholders, including those involved in the advocacy effort and those who are its targets. Case studies provide a full and detailed account about what happened. Isolating data points can, however, disguise the full story or context. Multiple case studies are useful when advocacy efforts are based in several locations or contexts, which allows comparisons across cases and can help in identifying patterns or existing and emerging themes (Coffman, 2010).

Endnotes

- 1** This is particularly relevant when using change-centred approaches to evaluation.
- 2** This same point is echoed in the context of evaluating peacebuilding, conflict transformation and aid in conflict settings. See, for example, Chigas, Church, Corlazzoli (2014); Church and Rogers (2006); Andersen, Bull, and Kennedy-Chouane (2014); Scharbatke-Church (2011).
- 3** Biases include selection bias (i.e. areas receiving humanitarian assistance are likely to have attributes that make them more or less likely to recover, compared to the average), and contamination bias (areas targeted by one actor are also likely to have other sources of assistance that may make it difficult to separate the different sources of changes) (Puri et al., 2014: v).
- 4** For the purpose of this Guide, advocacy is defined as any types of action or intervention that requires some form of influencing work (Tsui, Hearn, and Young, 2014:11).

Notes

Notes

Toolkits



Toolkit item #1: Evaluability checklists for the evaluation of protection

This guidance suggests that evaluability studies should cover four principal areas:¹

1. Overall level of ambition and types of question that evaluation stakeholders and programme stakeholders would like the evaluation to answer.
2. Programme design and intervention logic – particularly important for outcome and impact evaluations that make use of theory-based designs to understand causation, mixed-methods designs, and outcome-based approaches that look at contribution to results in multi-actor or networked interventions (e.g. outcome mapping, outcome harvesting).
3. Availability of data – or the possibility of generating additional data – required for the evaluation to answer the specific questions that commissioners and stakeholders have.
4. Conduciveness of the context to carry out an evaluation that looks at protection. This should include considerations regarding the organisational ‘climate’ and leadership support for the evaluation, considerations of access, logistics, safety of the evaluation team, and ethical appropriateness.

This toolkit consists of a set of four checklists covering these areas.

Evaluability checklist 1: Type of questions that evaluation and programme stakeholders would like to see answered

In the pre-evaluation stage, or during an evaluability study, there should be opportunities to get a sense of and clarify the expectations of different programme stakeholders and evaluation stakeholders regarding the evaluation exercise. This can be grasped by asking questions such as the ones in below.

- What are the sorts of question that programme stakeholders would like to see answered in an evaluative exercise or other reflective exercise? Do those questions relate to process and normative aspects of the intervention? Do the questions relate to cause-and-effect issues?
- Is there an expectation that the evaluation will focus mainly on issues and processes internal to the programme or intervention over which the agency should have more control and influence?
- Conversely, is the evaluation expected to look at a higher level of results (outcomes and impacts) that are beyond the sphere of control and (perhaps) influence of the intervention or programme and even of the agency itself, and that may touch on protection environment-wide issues and dynamics?
- Is the evaluation expected to cover process issues and results within the domain of humanitarian action, or to address multiple domains in the protection architecture, including development aspects (and global or country level – depending on the questions asked and the unit of analysis)?
- Is there broad alignment between the questions that programme stakeholders would like the evaluation to answer, and those that funders and other actors external to the programme would like to see answered? What are the implications for ensuring broad utility and for the accountability of the evaluation team itself?
- Are there realistic expectations for the evaluation to look at a higher level of results and environment-wide issues in terms of the level of resources, time, evaluation expertise, and protection know-how in the team that should carry out such exercise?

Evaluability checklist 2: Intervention logic/ programme design

A protection-oriented evaluability study should help uncover whether protection were – with any of its related domains of work and themes – incorporated in the earlier stages of the programme cycle (assessment, design, implementation and monitoring).

If so, it should be possible for the evaluators to discern how and to what extent protection issues had been spelt out in the protection analysis (if one had been carried out), in the frameworks used by the programme, and in related monitoring and reporting tools. (For a description of monitoring work and how it can provide different types of information along a result framework or logframe, see, for instance, Warner 2017; (IFRC, 2011b).

If, however, protection in humanitarian action had not been incorporated in the earlier stages of the programme cycle, this would translate into a need for greater efforts when gathering information to support the evaluative judgments (Faúndez and Weinstein, 2014: 11). The timing and scope of the evaluation should also be revised in light of those considerations.

During a pre-evaluation process, or during an evaluability assessment exercise, questions that can be asked might be formulated along the following lines:

- Does the programme clearly define the problem that it aims to change?
Is the expected change related to protection?
- If not, are there other references in the programme documents to 'do no harm' principles, to protection principles (see Sphere Project, 2012), or to other sectoral and thematic minimum standards for integrating and mainstreaming protection in humanitarian action (e.g. CPWG, 2012; Sutton et al., 2012)?
- Are the drivers of protection risks identified in the assessment, programme documents, or results framework?
- Has the expected beneficiary population of the programme been identified?
- Is the results framework of the programme coherently articulated? Do the outputs, outcomes and goal follow a coherent logic? How does protection feature in the resulting framework (e.g. as a set of specific activities with explicit result, or integrated in other sectoral interventions)?
- Are the objectives clear and realistic? Are they measurable (quantitatively or qualitatively)? Do they respond to the needs identified?
- Do proposed activities connect to the expected changes and desired results?

Evaluability checklist 3: Availability of information

During a pre-evaluation process, or during an evaluability assessment exercise, the type of questions might be formulated along the following lines:

Has the programme or intervention generated data needed to carry out disaggregated analysis by sex and age (at minimum), and by other characteristics, vulnerabilities, or other lines of affiliation to groups and sub-groups depending on the context and programme evaluated?

Was the initial programme or intervention design based on disaggregated data, and was this used to develop a protection analysis or other type of assessment and baseline studies?

Do project/implementing partners (if present) gather and use disaggregated information as part of monitoring day-to-day implementation and mid-course corrections during the life of the programme or project?

Are there gaps in the data generated by the programme? If so, is the evaluation expected to generate or reconstruct data to cover for those gaps in order to carry out the analysis and draw evaluative conclusions? Is it realistic to do so with available resources and within the timeframe of the evaluation?

Evaluability checklist 4: Conduciveness of the context

During a pre-evaluation process, or during an evaluability assessment exercise, the type of questions that can be asked could be formulated along the following lines:

- Would the internal conditions of the programme/project and the broader external conditions of the context within which the project is situated allow for an evaluation to take place? Are conditions conducive to ethical primary data collection and field visits?
- Are there resources, timing and security restrictions that should be taken into account at the scoping and design stage of the evaluation?
- Is there an adequate mix of skills and expertise in the programme ready to 'host' an evaluation mission?
- Are there sufficient human resources available at national/local level for the types of data collection that are to be undertaken? If not, is it possible to invest in developing the skills of the national/local evaluation team members who will undertake these tasks?

Figure 4 (in following page) presents an example developed by Cordula Reimann (2012) of a generic template for an evaluability checklist in the context of a peace-building initiative. Adaptation to a protection-oriented evaluation would of course require careful consideration on the part of the evaluation commissioners, but the structure presents a useful starting point for such work.

The checklist looks at the evaluability aspect of availability of information. Such checklist could easily be expanded to include more details around the elements expected to be in place to adjust along the spectrum from lower to higher evaluability.

Figure 4: Example evaluability checklists template looking at availability of information

Expected low evaluability conditions	Expected medium evaluability conditions	Expected high evaluability conditions
implicit ToC <input type="checkbox"/>	implicit but realistic ToC <input type="checkbox"/>	explicit ToC <input type="checkbox"/>
no baseline <input type="checkbox"/>	condensed baseline with data-gathering is focused on a few key indicators for selected goals <input type="checkbox"/>	complete baseline <input type="checkbox"/>
	no baseline but a more comprehensive monitoring at the beginning <input type="checkbox"/>	
no monitoring system <input type="checkbox"/>	monitoring system in place but not used routinely <input type="checkbox"/>	monitoring system in place to gather and systematise all necessary information <input type="checkbox"/>
	insights from monitoring are not translated into programme changes <input type="checkbox"/>	
no indicators <input type="checkbox"/>	indicators exist, but unrealistic, unmeasurable or unclear <input type="checkbox"/>	indicators exists and are SMART <input type="checkbox"/>
no access to stakeholders and programme participants or programme recipients <input type="checkbox"/>	difficult and limited access to stakeholders <input type="checkbox"/>	access to stakeholders <input type="checkbox"/>

Source: Adapted from Reimann (2012: 17)

Note: If most of the boxes ticked are in the left column, then the programme might not be ready for evaluation. In this case, it may be useful to reflect upon the full spectrum of available evaluative options. See [ALNAP \(2016\) EHA Guide](#) for more detail on these. It would also be worthwhile exploring with the programme team the paths to overcoming the weaknesses highlighted by the evaluability analysis.

Notes

Toolkit item #2: Evaluative rubrics

The key feature that sets evaluation apart from descriptive research is that evaluations require us to ask questions about how good something is, and whether it is good enough (Davidson, 2005). Evaluative rubrics are an increasingly common tool used to carry out this type of analysis in evaluation.

Evaluative rubrics are tables that describe what the evidence and indicators should look like at different levels of performance in order to make explicit how judgements are made in an evaluation when assessing the quality, value, or importance of an intervention or programme, policy or service provided. Originally developed and extensively used in the field of education evaluation, rubrics are made up of two main components:

1. the aspects of performance the evaluation focuses on (also referred to as ‘evaluative criteria’, ‘quality distinctions’, ‘merit criteria’, dimensions of merit or indicators)
2. descriptors that articulate what performance looks like at each level (also referred to as ‘merit determination’ (Scriven, 1991; Oakden, 2013: 5)

Why and how can rubrics be helpful to evaluators?

- They can help evaluators tackle the challenge of ‘valuing in evaluation’. This is about answering questions such as: on what basis do we make judgments about performance, quality, and effectiveness? And according to whom? (Julnes, 2012)
- They can help make transparent how the evaluators apply their professional judgment in order to draw succinct evaluative conclusions and for this reason they have been increasingly discussed in aid evaluation as a conduit for evaluative reasoning² (Davidson, 2005; 2014).
- They can be used as a ‘sense-making’ tool because ‘as the evidence layers and builds, it is possible to systematically make sense of many streams and lines of evidence, in a concise and cohesive way’ (King et al., 2013: 13).

Table 3: Example of ratings used to assess quantitative and qualitative data against each rubric

Rating	Quantitative and qualitative data
Excellent: Always	Clear example of exemplary performance or best practice in this domain; no weaknesses. Likely that 90% or more agree with statement to a considerable or high degree.
Very good: Almost always	Very good to excellent performance on virtually all aspects; scoring very well overall but not exemplary; no weaknesses of any real consequence. Possibly 80-90% agree with statement to a considerable or high degree.
Good: Mostly, with some exceptions	Reasonably good performance overall; might have a few slight weaknesses but nothing serious. In the range of 60-80% agree with statement to a considerable or high degree, and no more than 15% agree to a limited or very limited degree.
Adequate: Sometimes, with quite a few exceptions	Fair performance, some serious but non-fatal weaknesses on a few aspects. Around 40-60% agree with statement to a considerable or high degree, and no more than 15% agree to a limited or very limited degree.
Poor: Never (or occasionally, with clear weaknesses evident)	Clear evidence of unsatisfactory functioning; serious weaknesses across the board on crucial aspects. Probably less than 40% agree with statement to a considerable or high degree.

Source: Oakden (2013) originally adapted from Davidson (2005) and reproduced in Davidson (2014: 12)

How are data collected and analysed to populate evaluation rubrics?

There are two broad steps in developing rubrics:

1. The first step is to develop (usually in a participatory manner) the rich descriptions about the different (agreed) performance dimensions (indicators) and make explicit the different levels of performance of the programme or intervention.
2. The second is to consider the different types evidence (qualitative and quantitative) that might be used to draw a conclusion based on the definitions of performance.

A well-crafted rubric should paint the picture of what the mix of qualitative and quantitative evidence would look like, which also gives a clear sense of what will be needed to determine how performance should be rated.

Where existing data are to be used or the evidence has already been gathered, the key is not to define the rubric solely around what is available, but rather to paint the broad picture of what performance looks like regardless of existing evidence (Davidson, 2014: 6).

Examples of how rubrics have been used in evaluation can be found in Oakden (2013).

Notes

Toolkit item #3: A partial menu of evaluation approaches and designs

This toolkit item is a partial menu of evaluation approaches and designs that are of potential use for protection-oriented evaluations. It is not intended to be exhaustive but rather to present an initial overview of some of the options that evaluators may consider. Further options are presented and discussed in the ALNAP EHA Guide.

Participatory approaches

General features

Participatory evaluation approaches involve stakeholders in all aspects of the evaluation, including technical considerations.

The exercise of power and decision-making within the evaluation process itself shifts from the evaluator to the programme participants themselves. The evaluator's role shifts from expert to facilitator.

Patton (1997) described the basic principles of participatory evaluation as follows:

- Evaluation process involves participants' skills in goal-setting, establishing priorities, selecting questions, analysing data, and making decisions on the data.
- Participants own (commit to) the evaluation, as they make decisions and draw their own conclusions.
- Participants ensure that the evaluation focuses on methods and results they consider important.
- People work together, facilitating and promoting group unity.
- All aspects of the evaluation should be understandable and meaningful to participants.
- Facilitators act as resources for learning; participants act as decision-makers and evaluators.

Specific design applications and techniques

Empowerment evaluation (Fetterman, Kaftarian, Wandersman, 1996):

Empowerment evaluation aims to increase the probability of achieving programme success by providing programme stakeholders with tools for assessing the planning, implementation, and self-evaluation of their programme. This is often intended to lead to mainstreaming evaluation as part of the planning and management of the programme/organisation.

Action evaluation:

Action evaluation (based on concepts associated with action-research) is designed for stakeholders to develop and periodically refine meaningful programme goals and corresponding evaluation criteria throughout the course of their programme. It requires programme stakeholders to explicitly state and periodically revise their collective goals.

Through a series of self-reflection exercises stakeholders determine what they wish to achieve and what success will look like.

Utilisation-focused and developmental evaluation approaches

General features

A variety of methods and approaches to evaluation that focus explicitly on informing decision-making, helping organisations or groups to learn in real time and adapt their strategies to the changing circumstances around them.

Specific design applications and techniques

Patton's Development Evaluation (DE) (Patton, 2011):

- Development Evaluation (DE) is designed and facilitated to provide feedback, generate learning, and either supports strategy decisions or affirms changes to them.
- Choices about whether to use this approach should be based on judgements about the level of independence required in the evaluation and also the opportunities for engagement between evaluators and the programme over time.
- DE features internal and/or external evaluators who develop long-term relationships with programme participants.
- Evaluators become part of the programme team to ask evaluative questions, bring data and logic to the process, and facilitate evidence-based assessments and decision-making.
- Evaluators who are embedded may be viewed as having been less objective and neutral.
- Works well with: Complicated and complex strategies that evolve over time, and innovation and pilot initiatives in the development and testing phase.

Patton's Utilisation-focused evaluation approach (U-FE) (Patton, 2008):

U-FE is a process that can be structured following a 17-step process that starts with assessing and building programme and organisational readiness for U-FE to conclude with follow up with primary intended users to facilitate and enhance use, and meta-evaluation of use.

There is no specific content or method focus, and no specific methods of data collection and analysis.

Rather, U-FE adheres to a set of principles prescribing that the evaluation should be:

- judged by their utility and actual use
- situationally responsive
- negotiated process between evaluators, stakeholders and other evaluation users
- oriented towards facilitating decision-making about the issues being evaluated
- facilitated to support the involvement and engagement in the evaluation process and encourage uptake of evaluation findings.



Real-time evaluation (RTE) (Cosgrave, Ramalingam and Beck, 2009):

The principles underpinning RTEs in humanitarian action (which is where they are mostly commonly used) combine some features of DE and U-FE to ensure responsiveness to the fluid and fast-paced operational environments in which humanitarian actors work.

In an RTE, the primary objective is to provide feedback in a participatory way, during fieldwork, to those implementing and managing the humanitarian response.

Works well in the context of developing crisis, while response operations are ongoing, and when they are initiated early in an operation.

RTEs require evaluation team members not only evaluate what has been done but also to look at the plausible consequences of what is being done now. RTEs thus have both forward- and backward-looking components.

RTEs' primary stakeholders are the field team and those managing the operation from a head office. The evaluation team must communicate its findings to the team in the field, few of whom would have time to read a conventional evaluation report.

RTE reports should be finished or nearly finished by the time the team leaves the field.

Synthesis approaches

General features

One of their strengths is the ability to overcome some weaknesses of small sample sizes by compiling data from more than one study.

Key requirement: A strict coding protocol ensures consistency in interpretation. Poor coding protocols and coding errors are likely to undermine the validity of the study.

Specific design applications and techniques

Meta-analysis (Labin, 2008):

Meta-analysis is a quantitative tool that combines the results of different studies in order to yield new insights into the nuances surrounding results and changes at both outcome and impact level.

As a statistical method, meta-analysis requires the conversion of qualitative data into quantitative values.

One strength of meta-analysis is the ability to combine results across studies and samples to produce a better (more accurate, more statistically robust) estimate of the strength and stability of an intervention or of a relationship between two phenomena of interest.

Summary excerpt from: Corlazzoli and White (2013: 44)

Selected change-centred and theory-based approaches geared towards answering causal questions

Change-centred approaches to evaluation are geared to explore outcome- and impact-level results and changes, and deal with causal inference. They are intended to help in answering causal questions and establishing causal inference in evaluation.

The boxes that follow give an overview of evaluation approaches and possible design applications within the realm of change-centred approaches that aim to:

- Specify the basis on which the different methods and designs seek to infer causation (with different dominant orientations to establish attribution or contribution).
- Specify which approaches and specific designs and methods applications can work best in evaluations with small samples ('small n' – following the work by White and Phillips, 2012). 'Small n' evaluation scenarios are likely to be common in evaluating protection in humanitarian action, especially when purposive sampling is used.

Experimental designs

General features and specific design applications and techniques

Randomised Control Trials (RCTs) are often assumed to provide the strongest design option to control for selection bias in evaluation because the subjects are randomly assigned to the intervention.

They are most suitable for standardised interventions in identical settings with common beneficiaries and limited ranges of intervention modalities (limited variables to assess).

They infer causation based on random assignment of the intervention to groups.

They are less suitable for evaluations with complex causal factors or intentions to contribute to changes in the protection environment. In the 'messy' contexts of EHA protection, experimental designs will rarely be feasible or appropriate.



Quasi-experimental designs

General features

They make no use of randomisation. The evaluators 'construct' groups that are as equivalent on important characteristics (gender, income, socio-economic background) as possible.

Sometimes the evaluator can create a comparison group by matching key characteristics. For example, when large samples are used, or good secondary data is available, it is possible to use statistical matching techniques such as Propensity Score Matching.

Careful matching of treatment and comparison group can eliminate or greatly reduce the likelihood of rival explanations for a given result. An example of a rival explanation could be that the two groups were different from the start – and those different features are those that explain what brought about a result (Davidson, 2005: 246).

They are most suitable for standardised interventions in diverse settings, possibly with diverse beneficiaries

They infer causation based on establishing comparison groups and/or carrying out repeated measurement over time and/or carrying out before-and-after comparisons. As such, they require a relatively long evaluation timeframe or existence of strong and relevant monitoring data

Specific design applications and techniques

Regression discontinuity designs (RDD):

These are a powerful quasi-experimental statistical design where data are compared for a treatment and control group. People are not assigned randomly to groups but are chosen based on some cut-off value.

This makes this design suitable for use in humanitarian contexts as the treatment group can be selected by some value.

Although they are powerful, these designs work best when there are considerable amounts of data. The designs typically compare some dependent variable with the independent variable that is used as the cut-off for the treatment.

For example, if a programme targeted families with a particular household food security score, it would be possible to compare the score after six months against the original score, with separate regression lines for above and below the cut-off point.

Before-and-after designs without comparison group:

A before-and-after design gathers data at two points. The first is before starting an intervention. The second is after the intervention has begun. The goal of the design is to examine if the exposure has changed over time and infer whether this is connected to the intervention.

The point in time when the first measure ('before') and the second measure ('after') are taken varies. There is no standard rule on when this should be. It is not uncommon, though, to see time points six months to a year before and after an intervention.

Non-experimental of 'small n' evaluation approaches – Group I

General features

These involve evaluation approaches and designs that can be used to answer both descriptive and causal questions.

When used to answer causal questions they tend to focus on explaining the causal mechanisms at work in a given context.

They infer causation through the use of narrative and qualitative approaches to build plausible explanation of results.

Their goal is to explain what has occurred and how it has occurred.

Approaches below either seek out evidence to substantiate whether a programme's specified ToC was borne out in practice or they do the same for a number of alternative causal hypotheses which outline what might have occurred if causes or assumptions set out in the ToC had varied.

They attempt to establish contribution and causation beyond reasonable doubt by collecting evidence to validate, invalidate, or revise the hypothesised explanations, with the goal of documenting the links in the actual causal chain.

Specific design applications and techniques

Theory-Based Approaches to Evaluation (TBE) (Weiss, 2000; Funnell and Rogers, 2011):

TBE draw from and employ an explicit programme theory that: (1) spells out a set of hypothesised causal linkages between the intervention and desired outcomes provides; and (2) is used as a basis to analyse both attribution and contribution pathways.

If applied in a critically reflective manner, TBE can help distinguish poor theory from poor implementation (Weiss, 2000).

While the literature acknowledges that the findings may not be proven statistically, the approach can provide a logical argument that certain inputs will lead to a given change (Proudlock and Ramalingam, 2009) and should not necessarily be seen as a 'second best' option.

TBE should not be seen simply as a replacement for experimental and quasi-experimental designs. For high-stakes evaluations with large budgets and extended timelines, the two may be used in conjunction to strengthen causal attributions, provided they are used skilfully. For the everyday evaluator, working under time and budgetary constraints, ideas from both methodologies should be considered in order to build evidence for inferring causality (Cook, 2000, quoted in Davidson, 2000: 25).

Where a ToC is implicit or unarticulated, TBE may benefit from the participation of an external evaluator. Because TBE hinge on the clarity and strength of the theories of change, they are best served by evaluators with knowledge of the subject matter and TBE. TBE are resource-intensive and are most convincing when used in conjunction with other evaluation approaches such as Outcome Identification/measurement and Implementation Evaluation (Rogers, 2012).



No attempt is made to establish intervention and non-intervention groups and causation is inferred on the basis of:

- Identification/ confirmation of causal processes or 'chains'.
- Identification and confirmation of supporting factors and causal mechanisms at work in a given context.

Examples of specific design applications in TBE

- Contribution analysis (CA)

CA is an analytical tool using the intervention's strategic plan and assessing the contribution story. It is useful when there is no comparison group. It requires a strong ToC. (See [Toolkit items #4 and #6](#) for more step-by-step guidance.)

- Realist evaluation (RE) (Pawson and Tilley, 1997)

According to a realist perspective, programmes can be seen as theories incarnate; when a programme is implemented, it is testing a theory about what actions can help to bring about change (Westhorp et al., 2011).

Realist Evaluation sets out to establish: (1) an 'unequivocal causal relationship between a programme and outcome(s)'; and (2) that it was, beyond doubt, a programme that caused a given measurable change, and not some other factor(s). Programmes are viewed as being akin to open systems in which there are always multiple and competing mechanisms which interact with the surrounding context to produce outcomes. Pawson and Tilley (1997) sum this up as 'mechanisms + context = outcomes'.

All mechanisms interact with context, and so will not always achieve the same outcomes. RE is designed to explain how, and in what circumstances, programmes generate outcomes, by asking 'what works for whom, in what contexts, in what respects and how' (Pawson and Tilley, 1997).

An excellent summary of RE approaches produced by ODI Methods Lab is available in Westhorp (2014).

Case-based approaches:

They are most suitable for customised interventions in diverse settings with diverse beneficiaries that use narrative/qualitative approaches to build a plausible explanation of results

Infer causation based on comparisons across and within cases and analytical generalisation based on theory.

Examples of specific design applications in the case-based approaches group

- Qualitative Comparative Analysis (QCA) (Rihoux and Ragin, 2009)

QCA is an analytical tool used to compare multiple situations and determine different combinations of causal conditions.

This method is best used when there are multiple case studies with multiple factors to consider and when all factors are known. QCA will usually produce multiple 'causal recipes', relating to the different conjunctions of causal conditions, which produce a given outcome for a certain group of cases.

This technique is most suitable when several scenarios or aspects of an intervention need to be compared or understood. It can work also for 'medium n' evaluations.



- **Social Network Analysis (SNA) (Emirbayer and Goodwin, 1995)**

Social network analysis (SNA) is a methodology used to examine human behaviour and social change by analysing patterns of relations and relationships between individuals, groups, and/or organisations. SNA works to identify individuals or groups that have strong:

- Centrality: those with many relationships
- Prominence: those with the power and ability to influence networks and individuals
- Brokerage: those who can foster entrepreneurial relations or connections between others.

SNA views social relationships in terms of a 'network theory' made up of nodes (representing individual actors or groups within a network) with ties (representing the strength of the relationship or association with a line) (Emirbayer and Goodwin, 1995).

SNA can be used to measure social relationships in crisis, conflict and fragile environments. It can show who is connected to whom and the strength of the relationship within the larger network. It can also help in identifying who are the most significant actors or organisations that an intervention should target. It can also show which actors or organisations need support to be able to operate more effectively with others (Corlazzoli and White, 2013).

Non-experimental of 'small n' evaluation approaches – Group II

General features

Group II of small-n approaches have an explicitly participatory orientation when dealing with causal inference.

Approaches classified in this group are distinct from the Group I approaches because they do not set out to address attribution of cause and effect so explicitly.

In general, the Group II approaches place stakeholder participation at the heart of data collection and analysis. They focus on programme beneficiaries, implementers and other key stakeholders in order to establish what factors are perceived to have been important in producing change; in so doing, they aim to gain an insight into how a programme is performing and the part that it is playing in driving change (White and Phillips, 2012: 13).

Specific design applications and techniques

Most Significant Change (MSC) (Dart and Davies, 2003):

MSC is a form of participatory M&E that involves collecting stories at the field level and systematically analysing them to identify how stakeholders experience project outcomes and changes in the conflict.

MSC provides a method for capturing and analysing stories and exploring values behind the preferences for certain changes.

MSC may provide programme stakeholders and participants with a better understanding of what is and is not being achieved, and even whether they see achievements as valuable and relevant to them. Because of its open-ended questions, data can be collected about multiple dynamics or the project as a whole, rather than just the intended outcomes (Rogers, 2011).



Success Case Method (SCM) (Brinkerhoff, 2003; 2008):

SCM is a particular type of success case study that combines systematic and rigorous case-study methodology with storytelling, and reports results that stakeholders can easily understand and believe.

SCM is a narrative technique based upon naturalistic inquiry and in-depth case-study analysis. It is intended to be a quick and simple evaluation process geared towards understanding whether an initiative (such as a training or educational programme) is actually working.

SCM sets out to discover whether an intervention is working by searching for particularly successful or unsuccessful instances ('success' and 'non-success' cases).

SCM does not set out to find out about the 'average' participant, but instead intentionally seeks out the very best (and worst) that a programme has produced in order to understand the contribution that the programme has made to results, the role that contextual factors have played in influencing the different outcomes, and the way in which this information can be used to improve programme performance.

SCM is similar to other methods such as Appreciative Inquiry. (White and Phillips, 2012: 49)

Outcome Mapping (OM) (Earl et al., 2001; Smutylo, 2005; Ambrose and Roduner, 2009) (see also [Toolkit Item #4](#))

OM focuses on behavioural change and related outcomes such as capacity development and policy change.

The focus is on outcomes rather than the achievement of impacts, which are considered too 'downstream' in the results chain.

Rather than trying to accurately assess any one organisation's contribution to impact OM seeks to look at behaviours, resulting from multiple efforts, in order to help improve the performance of projects, programmes and policies.

With OM, 'boundary partners' – the individuals, groups and organisations that interact with projects, programmes and policies – are identified. OM assumes that the boundary partners control change more than the intervention itself.

The focus of OM is people. It represents a shift away from assessing the development impact of a project or programme towards describing changes in the way people behave through actions either individually or within groups or organisations.

OM provides a way to model what a programme intends to do. It differs from most traditional logic models because it recognises that different boundary partners operate within different logic and responsibility systems. OM can also be used as an end-of-programme assessment tool when the purpose of the evaluation is to study the programme as a whole.

OM proponents believe that many interventions, especially those focusing on capacity development, can better plan for and assess their contributions to development by focusing on behaviour (Morra Imas and Rist, 2009: 196-197).



OM proceeds through three stages:

1. Intentional design designates the intended macro-level changes and corresponding strategies.
2. Outcome and performance monitoring sets a self-assessment framework and data-collection tools for the ongoing monitoring of the programme's actions and progress towards results.
3. Evaluation planning sets the evaluation priorities and develops an evaluation plan.

OM recognises that multiple, non-linear events lead to change. OM looks at the logical links between interventions and behavioural change. OM assumes only that a contribution has been made, rather than assuming or attempting to claim attribution. (Rogers, 2011).

Participatory Impact Assessment (PIA) (Catley, Burns, Abede, and Suji, 2008):

PIA is an extension of Participatory Rural Appraisal (PRA) and involves the adaptation of participatory tools combined with more conventional statistical approaches specifically to measure the impact of humanitarian assistance and development projects on people's lives.

The PIA approach emphasises the standardisation and repetition of participatory methods, helping to improve the reliability of the information, but ideally leaving enough scope for the open-ended and flexible inquiry typical of PRA.

Can be used in both small-n and medium-n evaluations and sample sizes.

Well-designed PIA can assist communities and NGOs to measure impact using their own indicators and their own methods. PIA is designed around eight stages:

1. Defining the questions to be answered
2. Defining the boundaries of the project in space and time
3. Identifying and prioritising locally defined impact indicators
4. Deciding which method to use and testing it
5. Deciding which sampling method and sampling size to use
6. Assessing project attribution
7. Triangulation
8. Feedback and verifying results with the community

Sources: White and Phillips (2012); Stern (2008); Bamberger, Rugh and Mabry (2012); Stufflebeam and Coryn (2015); Tsui, Hearn and Young (2014); Mathison (2005); Rogers (2012); Morra Imas and Rist (2009); Chigas, Church and Corlazzoli (2014)

Notes

Toolkit item #4: ODI/RAPID approach to theories of change

The ODI-RAPID approach to developing theories of change follows the principles of Outcome Mapping (OM), which is used for both programming and M&E purposes focusing on behavioural and organisational change. Concretely, this means looking at changes in people's actions and behaviours within their organisations, not at changes in the things that are produced (Shaxson, 2014: 11).

The ODI/RAPID approach to OM-infused theories of change proceeds through three steps.

First step: Analyse the current context – this includes asking what ideas, interest groups and processes are influencing policy-making.

Second step: Examine for different stakeholders the **changes in actions and behaviours** that the agency:

- Expects to see: this indicates initial engagement with the intervention – early, positive responses to it.
- Would like to see: this indicates that there have been some initial changes (often called intermediate outcomes). At the level of behaviour change, this also indicates that key actors and programme recipients are showing signs that the messages are being taken on board and changing the way things are done.
- Would love to see: this indicates the higher-order changes towards which the intervention has been aiming. At the level of behaviour change, this indicates that the messages have been internalised.

Third step: Identify what the intervention will do, what others will do, and check assumptions about how these are related.

The lines between the different changes are often blurred, and it is often a matter of judgement regarding which change falls into which category (Shaxson, 2014: 11; Young, et al., 2014: 27).

Table 4: Example of an outcome-mapping -infused theory of change

	General statement of change	Which stakeholders are involved?	Specific indicators
Current context			
Expect to see: early positive response			
Like to see: active engagement			
Love to see: deep transformation in behaviour			

Note: More details on this tool are found in Young et al. (2014) and at www.roma.odi.org/further_resources_on_developing_an_engagement_strategy_to_influence_policy.html.

Toolkit item #5: ODI/RAPID influence and interest matrix

The RAPID team at the Overseas Development Institute (ODI) (www.odi.org/programmes/rapid) has developed a simple matrix (Figure 5) to map the stakeholders in policy and research influencing work that can also be used in assessing the ways that different actors in humanitarian protection can be influenced.

The matrix provides a conceptual basis for stakeholder-mapping efforts that can help evaluators to systematically assess what drives the interest, influence and actions of different stakeholders or explains their positions in a programme.

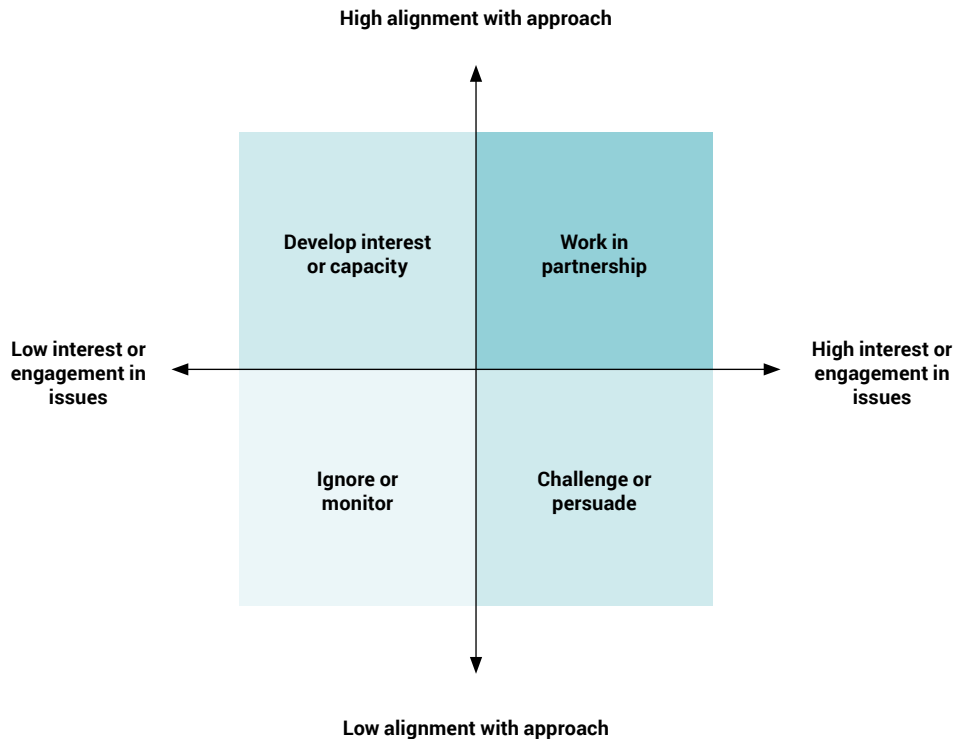
A large amount of useful information can be collected and put into a structured form to describe the relationships between different groups of people and how those groups are likely to behave when confronted with the possibility of change (Young et al., 2014: 14).

Compared to other stakeholder-mapping tools, this one is noteworthy because not only can it help to identify the main stakeholders in an intervention, but also suggest a possible course of action customised towards them (see Mendizabal, 2010 for a guide to using the matrix).

In the context of evaluating protection in humanitarian action, the tool can be used in evaluations that look at advocacy and behavioural change components in protection programming. It can also be useful when developing, customising and validating recommendations for different types of actors with varying degrees of interest, alignment and engagement with protection issues in a given context.

The tool can also be used at the analysis and design stage of an intervention. In the original ODI/RAPID formulation, the authors emphasise that ‘discussions about who is influential, why and what forms of interest they show in an issue can uncover important relationships between the stakeholders that you can subsequently use to develop your influencing objective. It will also make it more likely that you will consider the full range of people and organisations that need to be included’ (Young et al., 2014: 15).

Figure 5: The ODI/RAPID influence and interest matrix



Source: Adapted from Young et al., 2014

Toolkit item #6: Overview of contribution analysis

Contribution analysis (CA) involves six steps:

1. Develop the Theory of Change (ToC)
2. Assess the existing evidence on results
3. Assess alternative explanations
4. Assemble the performance story
5. Seek additional evidence
6. Revise and strengthen the performance story.

The steps are briefly described in Table 5. In essence, CA involves using evidence from existing assessment, monitoring and periodic evaluations to see what the data can reveal about the outcomes (or even impacts) of an intervention, while also considering what else besides the intervention could have brought about those results. A provisional performance story is developed from the existing data and should say something about:

1. The extent to which it is reasonable to assume that the programme/project's actions could have contributed to the observed outcomes.
2. The possible areas of weaknesses and where additional data would be useful.

Developing performance stories can be 'a powerful way of using existing data to determine what is known and where data is needed from additional forms of M&E, or if necessary from an impact evaluation, to provide a more convincing picture' (Perrin, 2013: 13).

Table 5: The six main steps in a contribution analysis process

Steps in contribution analysis	Description
<p>Step 1: Develop the results chain</p>	<p>Develop the programme theory model/programme logic/results chain describing how the programme is supposed to work. Identify the main external factors that might account for the outcomes observed. This programme theory should lead to a plausible association between the activities of the programme and the outcomes sought. Some links in the results chain will be fairly well understood or accepted. Others will be less well understood or subject to explanations other than that the programme was the 'cause'. In this way you acknowledge that attribution is indeed a problem.</p>
<p>Step 2: Assess the existing evidence on results</p>	<p>The results chain should provide a good idea of which intended results (outputs, intermediate and end outcomes) could be measured. What evidence (information from performance measures and evaluations) is currently available about the occurrence of these various results?</p> <p>The links in the results chain also need to be assessed. Which are strong (good evidence available, strong logic, or wide acceptance) and which are weak (little evidence available, weak logic, or little agreement among stakeholders)?</p>
<p>Step 3: Assess the alternative explanations</p>	<p>Outcomes are by definition influenced not only by the action of the programme but also by external factors – other programmes, as well as social and economic factors. In addition to assessing the existing evidence on results, there is a need to explicitly consider the extent of influence these external factors might have. Evidence or logical argument might suggest that some have only a small influence and that others may have a more significant influence on the intended results.</p>
<p>Step 4: Assemble the performance story</p>	<p>With this information, you will be able to set out your performance story of why it is reasonable to assume that the actions of the programme have contributed (in some fashion, which you may want to try and characterise) to the observed outcomes. How credible is the story? Do reasonable people agree with it? Does the pattern of results observed validate the results chain? Where are the main weaknesses in the story? There always will be weaknesses. These point to where additional data or information would be useful.</p> <p>If obtaining additional evidence is not possible (at least for now), then this is the most you can say about the extent to which the programme has made a difference.</p>
<p>Step 5: Seek out additional evidence</p>	<p>To improve your performance story you will need additional evidence. This could involve information on both the extent of occurrence of specific results in the results chain and the strength of certain links in the chain. A number of strengthening techniques that you might adopt are outlined in this work.</p>
<p>Step 6: Revise and strengthen the performance story</p>	<p>With the new evidence, you should be able to build a more credible story with which a reasonable person will be more likely to agree. It will probably not be fool proof, but will be stronger and more credible.</p>

Source: Mayne (2011) also cited in Perrin (2013)

Endnotes

- 1 The elements proposed could be used in general EHA work. They have been modified here but largely draw on, and are in line with, EHA guidance developed and piloted by various bilateral donor agencies (e.g. DFID, NORAD, Sida, USAID) and operational agencies including UNFPA, UNICEF and UN Women.
- 2 King et al. (2013: 20) went as far as arguing: 'We believe rubrics make evaluation accessible and create demand for evaluative thinking well beyond the group of people who think of themselves as evaluators.'

Notes



Overseas Development Institute
203 Blackfriars Road
London, SE1 8NJ

www.alnap.org/evaluation