



# Guide de l'évaluation de l'action humanitaire

**ALNAP** est un réseau d'apprentissage unique, composé de membres, dont l'objectif est d'améliorer la qualité et la redevabilité de l'action humanitaire.

**[www.alnap.org](http://www.alnap.org)**

**Auteurs :**

Margie Buchanan-Smith

John Cosgrave

Alexandra Warner

**Avec le concours de :**

Francesca Bonino

Neil Dillon

Paul Knox Clarke

Ralf Otto

Groupe URD (traduction : Pierre Brunet ; relecture :

Jean-Christophe Brard et Salomé Cousinié)

IECAH (révision: Camille Nussbaum)

Les opinions exprimées dans ce rapport ne reflètent pas nécessairement celles des membres d'ALNAP.



ALNAP (2016). Guide de l'évaluation de l'action humanitaire. Guide ALNAP. Londres : ALNAP/ODI.

© ALNAP/ODI 2016. Cette œuvre est sous licence Creative Commons Attribution-Non commerciale (CC BY-NC 4.0).

ISBN 978-1-910454-49-7

Responsables de la communication :

Maria Gili et Alex Glynn

Conception : Human After All

[www.humanafterall.co.uk](http://www.humanafterall.co.uk)

## Avant-propos

par John Mitchell



Lorsqu'ALNAP a vu le jour il y a vingt ans, l'évaluation humanitaire en était encore à ses premiers pas. Depuis, l'EAH est devenue un élément important du fonctionnement du système humanitaire et l'on ne compte plus les évaluations de qualité. De ce fait, l'apprentissage et la redevabilité se sont fortement améliorés.

Les connaissances tirées de la pratique de l'EAH constituent désormais une base mondiale de données probantes qui ne cesse d'augmenter. Celle-ci est utilisée pour créer des seuils de référence à partir desquels la « performance » humanitaire est suivie dans le temps, comme le met en évidence le rapport *L'état du système humanitaire* publié par ALNAP. Le besoin de matériel évaluatif de grande qualité pour améliorer la précision et le niveau de confiance associés aux informations portant sur la performance sera toujours vital pour la communauté humanitaire. A cet égard, je pense que ce Guide sera d'une grande utilité.

Concevoir un Guide vraiment complet a été une tâche colossale qui a nécessité cinq années. Cet ouvrage couvre de nombreux aspects de l'évaluation humanitaire, la plupart s'inscrivant dans un environnement opérationnel particulièrement complexe. Ce Guide a également bénéficié d'un taux de participation inhabituellement élevé avec 40 organisations qui l'ont testé et commenté lors du processus pilote. Sans ces contributions, nous n'aurions pas pu créer un guide aussi clair et pertinent qui fait autorité et a déjà prouvé sa valeur inestimable aux évaluateurs. Aussi, j'aimerais remercier sincèrement tous ceux qui y ont contribué.

Nous espérons que ce Guide sera largement utilisé et contribuera à la qualité et au plaisir de réaliser des évaluations dans le domaine de l'action humanitaire. Il continuera également à renforcer la redevabilité et à promouvoir de meilleures données évaluatives en faveur de la capitalisation et de l'analyse, que ce soit au niveau de l'ensemble du système, de chaque organisation ou de leurs programmes. Nous invitons enfin les utilisateurs de ce Guide à continuer de nous faire part de leurs retours puisque nous tenons à garantir la qualité des orientations exposées dans cet ouvrage et, par conséquent, l'impact de l'EAH.



# Table des matières

## Introduction

03	Avant-propos
10	Liste des acronymes
12	Remerciements
14	Pourquoi l'évaluation est importante
14	Pourquoi un Guide de l'évaluation de l'action humanitaire ?
16	L'élaboration du Guide ALNAP d'EAH
18	Comment utiliser le Guide EAH d'ALNAP
20	Comment ce Guide est-il organisé ?
22	Légende des éléments graphiques

## 1 / Qu'est-ce que l'évaluation de l'action humanitaire ?

24	1.1 Crises humanitaires et action humanitaire
26	1.2 Evaluation de l'action humanitaire (EAH)
30	1.3 Suivi et évaluation dans le cycle de réponse humanitaire
32	1.4 Les défis les plus courants de l'EAH

## 2 / Décider de réaliser une évaluation

41	2.1 Quand l'évaluation convient-elle à des objectifs de redevabilité et d'apprentissage ?
47	2.2 L'évaluation est-elle le bon outil pour le travail demandé ?
48	2.3 Equilibrer apprentissage et redevabilité
49	2.4 Décider quoi et quand évaluer
55	2.5 Ethique et EAH

## 3 / Penser tôt et souvent à l'utilisation de l'évaluation

62	3.1 Ce que signifie se focaliser sur l'utilisation
63	3.2 Comment sont utilisées les évaluations de l'action humanitaire
65	3.3 Qui sont les utilisateurs visés de l'évaluation ? Quels sont leurs besoins en information et comment peuvent-ils être impliqués ?
72	3.4 Le rôle de la population affectée dans la phase de planification
73	3.5 Les facteurs affectant l'utilisation des évaluations

## 4 / Les types d'évaluation

- 78 4.1 Quel type d'évaluation allez-vous entreprendre ?
- 79 4.2 Définir les différents types d'évaluation
- 87 4.3 Les évaluations en temps réel (ETR)
- 89 4.4 Les évaluations conjointes

## 5 / Elaborer votre évaluation

- 92 5.1 La logique du programme
- 97 5.2 Les théories du changement
- 100 5.3 Développer un modèle logique a posteriori
- 101 5.4 Les cadres d'évaluation et les critères d'évaluation

## 6 / Choisir les questions d'évaluation

- 104 6.1 Le choix des questions d'évaluation a un effet critique sur la qualité
- 107 6.2 Les types de questions d'évaluation
- 109 6.3 Réduire le nombre de questions
- 110 6.4 Décomposer les grandes questions
- 111 6.5 Les critères d'évaluation
- 116 6.6 Choisir les meilleures questions

## 7 / Les termes de référence et la budgétisation d'une évaluation

- 118 7.1 Termes de référence (TdR)
- 119 7.1.1 Les TdR et le rapport de démarrage définissent l'évaluation
- 121 7.1.2 Que trouve-t-on dans les TdR ?
- 127 7.1.3 Le calendrier des TdR
- 127 7.1.4 La longueur des TdR
- 128 7.2 La budgétisation d'une évaluation

## 8 / La phase de démarrage

- 136 8.1 Le rapport de démarrage et le contrôle qualité
- 138 8.2 Les activités de la phase de démarrage
- 138 8.3 La matrice d'évaluation
- 142 8.4 Le rapport de démarrage

## 9 / Planifier et gérer votre évaluation

- 147 9.1 La décision d'évaluer
- 148 9.2 Les groupes consultatifs
- 153 9.3 Le calendrier de l'évaluation
- 154 9.4 Les évaluations internes et externes
- 157 9.5 Recruter les évaluateurs
- 168 9.6 Les défis liés au *leadership* et au travail en équipe
- 170 9.7 Gérer les conflits

## 10 / Les méthodes de recherche documentaire

- 172 10.1 Qu'est-ce qu'une étude documentaire ?
- 174 10.2 Pourquoi réaliser une revue documentaire ?
- 175 10.3 Combien de temps doit durer une revue documentaire ?
- 175 10.4 Réaliser une recherche documentaire
- 178 10.5 Les outils de la recherche documentaire
- 182 10.6 Les produits issus de la revue documentaire

## 11 / Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation

- 193 11.1 Qualitatives et quantitatives
- 194 11.2 Les familles de conception
- 200 11.3 Choisir une conception
- 209 11.4 Les biais

## 12 / L'échantillonnage

- 215 12.1 Les problèmes d'échantillonnage courants dans l'EAH
- 216 12.2 L'échantillonnage non aléatoire
- 224 12.3 L'échantillonnage aléatoire
- 230 12.4 L'échantillon pour les méthodes combinées

## 13 / Les méthodes de terrain

- 232 13.1 La technique de l'entretien
- 236 13.2 Les interprètes
- 238 13.3 Les méthodes d'entretien qualitatif
- 246 13.4 Les méthodes de sondage
- 253 13.5 L'observation
- 257 13.6 Les mesures discrètes
- 259 13.7 Les méthodes tournées vers l'apprentissage

## 14 / Impliquer la population affectée dans votre évaluation

- 268 14.1 Pourquoi doit-on impliquer la population affectée ?
- 271 14.2 Planifier l'implication avec la population affectée
- 273 14.3 Concevoir l'évaluation pour impliquer la population affectée
- 274 14.4 Les méthodes d'implication de la population affectée
- 279 14.5 Les difficultés propres à l'implication de la population affectée dans l'EAH
- 280 14.6 Les questions éthiques liées à l'implication de la population affectée

## 15 / L'accès restreint

- 283 15.1 Qu'entendons-nous par accès restreint ?
- 284 15.2 L'importance d'une étude de faisabilité de l'évaluation
- 286 15.3 Comment surmonter l'accès restreint ?
- 291 15.4 La crédibilité de l'évaluation à distance
- 292 15.5 Les autres options à prendre en compte si une évaluation à distance est irréalisable



## 16 / L'analyse

- 294 16.1 Analyse quantitative ou qualitative
- 296 16.2 Analyser les éléments de preuve pour répondre à des questions normatives
- 297 16.3 Analyser les éléments de preuve pour des questions évaluatives
- 302 16.4 Analyser les éléments de preuve pour des questions causales
- 303 16.5 Analyser les éléments de preuve pour des questions descriptives
- 311 16.6 L'analyse statistique des données primaires
- 316 16.7 L'analyse numérique des données secondaires, comme celles issues des registres de données de distributions
- 318 16.8 Des éléments de preuve aux recommandations

## 17 / Rédiger et communiquer des résultats d'évaluation en se focalisant sur l'utilisation

- 328 17.1 Les produits clés
- 330 17.2 Le rapport d'évaluation
- 339 17.3 La diffusion
- 342 17.4 Les différents canaux de communication
- 346 17.5 Faciliter l'adoption d'une évaluation
- 349 17.6 Les synthèses d'évaluation, les revues thématiques et les méta-analyses

## 18 / Les évaluations d'impact humanitaire

- 356 18.1 Pourquoi une évaluation d'impact est-elle importante ?
  - 356 18.2 Qu'est-ce que l'impact dans l'EAH ?
  - 359 18.3 Planifier une évaluation d'impact de l'action humanitaire
  - 362 18.4 Approches et conceptions
- 
- 367 Glossaire

La bibliographie et le sommaire sont disponibles à l'adresse suivante :  
[www.alnap.org/EHA](http://www.alnap.org/EHA)

## Liste des acronymes

<b>ACF</b>	Action Contre la Faim
<b>ACT</b>	Action by Churches Together
<b>ALNAP</b>	Active Learning Network for Accountability and Performance in Humanitarian Action (Réseau actif d'apprentissage pour la redevabilité et la performance dans l'action humanitaire)
<b>CAQDAS</b>	Computer Assisted Qualitative Data Analysis (analyse des données qualitatives assistée par ordinateur)
<b>CDAC</b>	Communicating with Disaster Affected Communities (réseau Communiquer avec les communautés affectées par une catastrophe)
<b>CDA</b>	CDA Collaborative Learning Projects (projets d'apprentissage collectif)
<b>CERF</b>	Central Emergency Response Fund (Fonds central d'intervention d'urgence)
<b>CFW</b>	Cash for work (argent-contre-travail)
<b>CHS</b>	Core Humanitarian Standard (Norme humanitaire fondamentale)
<b>CICR</b>	Comité international de la Croix-Rouge et du Croissant-Rouge
<b>CRS</b>	Catholic Relief Services
<b>CwC</b>	Communicating with Communities (Communiquer avec les communautés)
<b>DEC</b>	Disasters Emergency Committee
<b>DFID</b>	UK Department for International Development (département du gouvernement britannique en charge du développement international)
<b>DRC</b>	Danish Refugee Council
<b>ECB</b>	Emergency Capacity Building Project
<b>ECHO</b>	Direction générale pour la protection civile et les opérations d'aide humanitaire (Commission européenne)
<b>EAH</b>	Evaluation de l'action humanitaire
<b>EPR</b>	Evaluation participative rapide
<b>ETR</b>	Evaluation en temps réel
<b>FAO</b>	Organisation des Nations unies pour l'alimentation et l'agriculture
<b>FICR</b>	Fédération internationale des Sociétés de la Croix-Rouge et du Croissant-Rouge
<b>FFS</b>	Farmer Field School
<b>GNUE</b>	Groupe des Nations unies pour l'évaluation
<b>HAP</b>	Humanitarian Accountability Partnership
<b>IASC</b>	Inter-Agency Standing Committee (Comité permanent inter-organisations)
<b>IRC</b>	International Rescue Committee

<b>JEEAR</b>	Joint Evaluation of Emergency Assistance to Rwanda (évaluation conjointe de l'aide d'urgence au Rwanda)
<b>JHIE</b>	Joint Humanitarian Impact Evaluation (évaluation conjointe d'impact humanitaire)
<b>LRRD</b>	Linking Relief Rehabilitation and Development (liens entre urgence, réhabilitation et développement)
<b>MEAL</b>	Monitoring, evaluation and learning (suivi, évaluation et apprentissage)
<b>MSF</b>	Médecins Sans Frontières
<b>NRC</b>	Norwegian Refugee Council
<b>OCHA</b>	United Nations Office for the Coordination of Humanitarian Affairs (Bureau de la coordination des affaires humanitaires)
<b>ODI</b>	Overseas Development Institute
<b>OCDE-CAD</b>	Organisation de Coopération et de Développement Économiques – Comité d'aide au développement
<b>OFDA</b>	Office of US Foreign Disaster Assistance (USAID)
<b>OMS</b>	Organisation mondiale de la santé
<b>ONG</b>	Organisation non gouvernementale
<b>PAM</b>	Programme alimentaire mondial
<b>PDI</b>	Personnes déplacées internes
<b>PNUD</b>	Programme des Nations unies pour le développement
<b>RAPID</b>	Overseas Development Institute's Research and Policy in Development Programme
<b>RCA</b>	République centrafricaine
<b>RCT</b>	Randomised control trial (essais contrôlés randomisés)
<b>RDC</b>	République démocratique du Congo
<b>ROI</b>	Region of Origin Initiative in Afghanistan
<b>ROMA</b>	RAPID Outcome Mapping Approach
<b>RRC</b>	Réduction des risques de catastrophes
<b>S&amp;E</b>	Suivi et évaluation
<b>SCHR</b>	Steering Committee for Humanitarian Response
<b>Sida</b>	Swedish International Development Cooperation Agency (Agence suédoise de coopération internationale au développement)
<b>TEC</b>	Tsunami Evaluation Coalition
<b>TdC</b>	Théorie du changement
<b>TdR</b>	Termes de référence
<b>UNHCR</b>	Haut-Commissariat des Nations unies pour les réfugiés
<b>UNICEF</b>	Fonds des Nations unies pour l'enfance
<b>WASH</b>	Water, sanitation, and hygiene (eau, assainissement et hygiène)
<b>WVI</b>	World Vision International

## Remerciements

Processus à la fois vaste et participatif, l'élaboration du Guide de l'évaluation de l'action humanitaire a été rendu possible grâce à la disponibilité, à l'énorme travail et au dévouement de nombreux individus et organisations. Le Secrétariat d'ALNAP aimerait en remercier certains en particulier.

Tout d'abord, ce Guide n'aurait pas été possible sans l'implication toujours positive des membres d'ALNAP. Le Secrétariat d'ALNAP est extrêmement reconnaissant envers le réseau pour son honnêteté et son enthousiasme : il n'aurait pas été possible de créer un tel outil pratique sans le soutien des membres.

Le Secrétariat d'ALNAP souhaiterait également remercier John Cosgrave et Margie Buchanan-Smith pour leur engagement constant et leur travail passionné pour rédiger, améliorer, promouvoir et utiliser ce Guide EAH ainsi que les précédents supports de formation EAH des dernières années.

Nous adressons aussi un remerciement très spécial à Wendy Fenton, du Humanitarian Practice Network (HPN), qui a lancé une enquête au sein des membres du HPN. La plupart d'entre eux a reconnu le besoin d'un nouvel outil portant sur l'évaluation de l'action humanitaire.

Pour leurs précieux retours, notre reconnaissance va également aux membres du groupe consultatif représentant les parties prenantes des évaluations au sein du secteur : Jock Baker, Tony Beck, Hana Crowe, Stefan Dahlgren, Wendy Fenton, Enrique Garcia, Josse Gillijns, Saul Guerrero, Babar Kabir, Peter Klansoe, Caroline Loftus, Rob McCouch, Joakim Molander, Jonathan Patrick, Nicoletta Pergolizzi, Riccardo Polastro, Hannah Reichardt, Peta Sandison et Vivien Walden. Ils ont permis de guider le Secrétariat d'ALNAP tout au long du processus de rédaction.

Nous adressons aussi un immense merci aux centaines de personnes qui ont testé et utilisé ce guide EAH, prenant le temps de s'impliquer dans cette initiative et de revenir vers nous avec des éclairages, des commentaires, des suggestions ou simplement un mot de reconnaissance et de gratitude. Nous avons été émerveillés par la volonté des membres du réseau ALNAP, et au-delà, de prendre part à la phase de pilotage, de réviser le module de *e-learning* co-développé avec UNICEF et de partager leurs points de vue quant à la manière de rendre ce Guide le plus utile possible et focalisé sur le praticien. Leurs retours ont été d'une valeur inestimable dans la révision et la finalisation de ce Guide EAH.

Merci également au groupe restreint de testeurs et de praticiens EAH qui nous a aidés à traiter les retours les plus difficiles et les plus pointus durant l'atelier de validation des retours organisé à la fin de la phase pilote. Merci donc à Sarah Bailey,

Lian Bradley, Annie Devonport, Kathy Duryee, Josse Gillijns, Langdon Greenhalgh, H elene Juillard, Anne-Claire Luzot, Velina Mikova, Mikkel Nedergaard, Joanna Olsen, Juliet Parker, Koorosh Raffii, Anke Reiffenstuel et Bonaventure Sokpoh. Nous sommes aussi reconnaissants   Ralf Otto qui a anim  cet  change et repr sent  ce groupe en tant que *peer-reviewer* pour la version finale.

Nous envoyons  galement un chaleureux remerciement   Sian Cook qui a fourni un appui   la recherche de grande valeur durant la r vision de ce Guide EAH.

Un grand merci   John Mitchell et Paul Knox Clarke pour leur encadrement et leurs conseils de chaque instant, ainsi qu'  l' quipe Communication du Secr tariat ALNAP pour son soutien et sa cr ativit  de tous les instants qui ont permis de faire de ce Guide EAH un document accessible et agr able. Nos remerciements vont en particulier   Patricia Curmi, Franziska Orphal, Maria Gili et Alex Glynn.

Notre dernier remerciement, mais pas le moindre, est adress    Francesca Bonino qui a coordonn  cette initiative pendant quatre ann es. Francesca, ton  nergie et ton enthousiasme ont aid    faire avancer ce projet qui n'aurait pas  t  possible sans tes efforts.

## Pourquoi l'évaluation est importante

Comment savons-nous que nos efforts humanitaires sont des réussites ? L'évaluation est l'un des principaux moyens permettant de le découvrir. Fondamentalement, l'évaluation vise à produire un jugement éclairé sur la valeur d'activités et leurs résultats. Avons-nous pesé de façon positive sur la situation ? Avons-nous vraiment aidé à sauver des vies et atténuer les souffrances ? Avons-nous accompli cela de la meilleure façon possible ? Des évaluations sérieuses nous permettent de porter un jugement réfléchi fondé sur des données probantes par rapport au degré de réussite d'un programme, et la nature précise de cette réussite. Elles permettent aux parties prenantes de répondre à la question « Et alors ? » (Morra Imas et Rist, 2009), en regardant au-delà des généralités formulées sur cette réussite pour se focaliser sur les raisons pour lesquelles un projet, un programme ou une politique a été utile – ou non - et pour qui (*Buffardi et al.*, 2015).

L'effort impulsé dans les années 90 pour améliorer la qualité et la redevabilité du secteur humanitaire (Knox-Clarke et Darcy, 2013), ainsi que les pressions récentes pour démontrer le bon rapport qualité-prix des interventions et optimiser les ressources devenues de plus en plus rares (ALNAP, 2015), ont encore accentué l'importance des évaluations de l'action humanitaire. Les évaluations bien planifiées, conçues et réalisées sont des processus susceptibles d'aider à la fois à l'apprentissage et à la redevabilité, et ce à plusieurs niveaux. Dans ces contextes de travail au rythme souvent rapide, les évaluations constituent des opportunités intéressantes pour marquer une pause et prendre du recul, en adoptant un point de vue objectif sur ce qui fonctionne ou a fonctionné, et ce qui doit changer, en apportant un éclairage structuré sur la performance globale d'un projet, d'un programme ou d'une réponse. Elles contribuent à l'ensemble de données probantes relatives à ce qui « fonctionne » ou non dans ces contextes souvent complexes. Ainsi, les évaluations peuvent répondre à la difficile question « Quels sont nos véritables résultats ? » et aider les décideurs à faire les réorientations nécessaires ou à effectuer des choix difficiles.

## Pourquoi un Guide de l'évaluation de l'action humanitaire ?

Les évaluations de l'action humanitaire ont connu une forte augmentation depuis les années 90. Cela a donné lieu à la création d'une masse critique de connaissances collectives et permis d'obtenir une expérience sur lesquelles s'appuyer. Jusqu'en 2009, il n'existait toutefois aucun guide détaillé pour le secteur en ce qui concerne l'évaluation des interventions humanitaires. Le travail sur ce Guide a débuté en réponse à des retours de membres du Humanitarian Practice Network : lorsqu'on leur a demandé quelle serait leur principale requête en matière de document d'orientation, l'EAH fut clairement la réponse majoritaire.

## Six raisons montrent que le moment est venu de proposer un Guide EAH détaillé :

1

L'intérêt et les investissements augmentent en ce qui concerne les évaluations car des inquiétudes sont apparues au niveau de la redevabilité et de l'efficacité du développement international et de l'action humanitaire.

2

Il existe à présent une masse critique de connaissances collectives sur laquelle s'appuyer – la seule base de données des rapports d'évaluation d'ALNAP contient plus de 2.000 évaluations couvrant les trente dernières années.

3

On constate un besoin de créer un langage commun et une compréhension partagée de l'EAH de manière à faciliter les échanges au sein des équipes et des organisations, mais aussi entre organisations.

4

Même si les évaluations sont devenues une pratique plus courante, seules quelques-unes font preuve d'une méthodologie claire. Et lorsque c'est le cas, les organisations humanitaires et les évaluateurs s'en tiennent souvent à un nombre restreint de type de conceptions et de méthodes. Or, il est toujours possible d'améliorer la gamme des conceptions et des méthodes utilisées ainsi que le choix de la méthodologie la plus efficace pour les questions d'évaluation qui sont posées.

5

De même, les évaluations sont toujours un outil commun permettant aux bailleurs de juger la redevabilité, mais elles peuvent également être utilisées par les organisations afin de comprendre et d'améliorer leurs programmes. De façon générale, il existe encore de nombreuses opportunités pour que les organisations humanitaires tirent le maximum de profit des évaluations qui correspondent à ces besoins.

6

La commande des évaluations est passée des sièges des organisations au personnel sur le terrain car les organisations sont en cours de décentralisation. Pour autant, les responsables sur le terrain ont souvent peu d'expérience en matière de planification et de gestion des évaluations, en particulier des EAH.

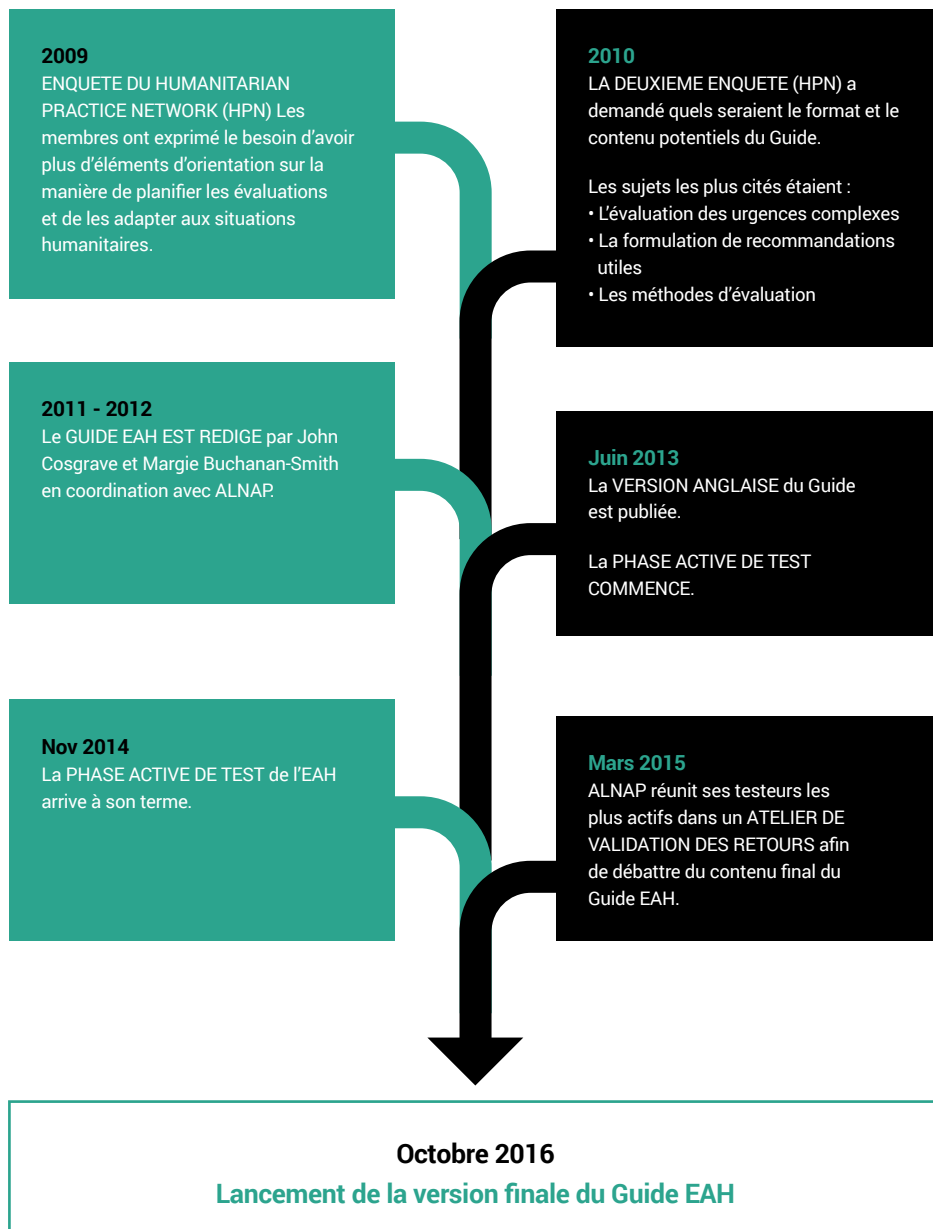
## L'élaboration du Guide ALNAP d'EAH

L'élaboration du Guide ALNAP de l'évaluation de l'action humanitaire s'est appuyée sur de vastes recherches et consultations. Ce Guide a été :

- Conçu à partir de supports de formation commandités par ALNAP en 2001. Ces documents ont été testés, mis à jour et adaptés de nombreuses fois au fil des années.
- Rédigé en consultation avec un groupe consultatif qui représentait les parties prenantes de l'évaluation pour l'ensemble du secteur.
- Testé de façon active sur une période de 18 mois. Plus de 40 organisations humanitaires ont utilisé ce Guide et fourni des retours. Ce processus pilote a abouti à un atelier de validation des retours hébergé par ALNAP, avec 15 des testeurs et praticiens de l'EAH les plus actifs. Ce groupe a aidé le Secrétariat d'ALNAP à décider comment prendre en compte les retours les plus complexes lors du processus de révision finale.
- Revu et mis à jour durant la conception de l'introduction ALNAP-UNICEF au module de *e-learning* sur l'EAH, en collaboration avec UNEG et EvalPartners. Les sections révisées de ce Guide se sont fortement inspirées du processus de révision complète qui a eu lieu pour le repenser dans un format en ligne.
- Révisé en s'appuyant sur les retours reçus pendant le processus de test. Un examen des rapports d'évaluation et d'autres ressources évaluatives a été utilisé pour intégrer de nouveaux exemples.



## Animation de l'historique de l'EAH



# Comment utiliser le Guide EAH d'ALNAP

## A qui s'adresse ce Guide ?

Ce Guide est destiné à aider toutes les personnes qui réfléchissent à la planification, la conception et la mise en œuvre des évaluations et/ou cherchent à les utiliser, en s'appuyant sur la masse critique de connaissance, et en particulier sur un grand nombre d'exemples de bonnes pratiques. Ce Guide vise à satisfaire un besoin exprimé, ainsi qu'à aider les commanditaires, les responsables, les concepteurs, les chargés de mise en œuvre et les utilisateurs d'évaluations à tirer le meilleur parti du processus d'évaluation et, *in fine*, à améliorer la qualité des EAH et la manière dont elles sont utilisées.

## Éléments à prendre en compte au moment d'utiliser ce Guide

Ce Guide vise à appuyer les évaluations de qualité qui contribuent à de meilleurs résultats en fournissant les meilleures données probantes disponibles sur ce qui fonctionne bien, ou non, et pourquoi. Dans ce but, les lecteurs doivent garder à l'esprit quelques éléments lorsqu'ils utilisent ce Guide.

Avant toute chose, les évaluations de l'action humanitaire coûtent de l'argent qui aurait pu être utilisé pour éviter des décès ou réduire des souffrances. Cet argent est bien dépensé s'il amène des améliorations au sein de l'action humanitaire, mais cela ne sera le cas que si les conclusions des évaluations sont de qualité et utilisées (Hallam et Bonino, 2013). Une évaluation ne doit pas être une activité « autonome », mais faire partie des efforts réalisés par une organisation pour rendre des comptes sur son travail et pour apprendre. Les évaluations sont susceptibles d'avoir leur plus grande influence en matière de contribution à l'amélioration des performances lorsque les organisations, et le secteur dans son ensemble, sont investis dans la compréhension de leur manière de travailler et dans l'apprentissage.

Par conséquent, l'EAH peut faire beaucoup plus que remplir une simple fonction de redevabilité. Cela ne signifie pas que les évaluations ne sont pas un outil important pour la redevabilité mais la redevabilité ne se limite pas à de simples rapports. Elle relève aussi de l'internalisation du contenu de ces rapports et des actions mises en œuvre pour réaliser des améliorations à partir de ces données. L'apprentissage en est un élément important et comme l'affirme Irene Guijt : « Vous ne pouvez pas être redevable si vous n'apprenez pas. » (2010: 277). Très souvent, les évaluations sont perçues, voire redoutées, comme une forme de critique ou seulement comme une exigence du bailleur. Pourtant, elles peuvent – et doivent – jouer un rôle beaucoup plus constructif et utile. En adaptant les évaluations aux besoins de leurs principaux utilisateurs visés, elles constituent des opportunités exceptionnelles pour l'apprentissage et l'amélioration des programmes. Ce Guide a été écrit en se focalisant fortement sur l'utilisation.

La plupart des défis et des difficultés rencontrés par les personnes impliquées dans l'action humanitaire affectent également les évaluations de l'action humanitaire, le manque d'accès aux populations affectées et aux zones touchées étant peut-être l'une des caractéristiques les plus complexes (Hallam et Bonino, 2013: 12). Ainsi, comment pouvons-nous mener des évaluations suffisamment rigoureuses et crédibles dans de tels contextes ? Parce qu'elle se trouve au cœur de ce Guide, la réponse à cette question est développée dans chaque section.

Le Guide EAH présente une série d'étapes structurées pour vous aider à atteindre des résultats de qualité. Chacune de ces étapes doit être adaptée au contexte de l'évaluation (en prenant en compte des facteurs comme : la crise, la réponse, le pays, la région, le projet ou programme, ou l'équipe) ainsi qu'au contexte organisationnel (par exemple, l'« appropriation » par la direction, la culture de l'apprentissage). Ce Guide présente des exemples d'évaluations de grande et de petite échelle. Il est important de définir le niveau d'ambition d'une EAH en fonction des ressources disponibles, c'est-à-dire en termes de capacités internes et de fonds. Pour vous y aider, ce Guide EAH offre certains éclairages sur les évaluations de plus petite dimension.

Ce Guide s'appuie sur un certain nombre de publications et de ressources d'ALNAP sans faire pour autant doublon. Même s'il vise à être complet, il est très probable que certains sujets actuels ne sont pas traités et que d'autres sont susceptibles d'émerger puisque l'action humanitaire évolue sans cesse. On peut citer comme exemple l'évaluation des activités de protection humanitaire. Le réseau ALNAP prépare actuellement un guide sur l'évaluation de la protection qui sera publié en complément de ce Guide. Au fil du temps, d'autres sujets peuvent émerger et être traités dans des guides annexes.

## Comment ce Guide est-il organisé ?

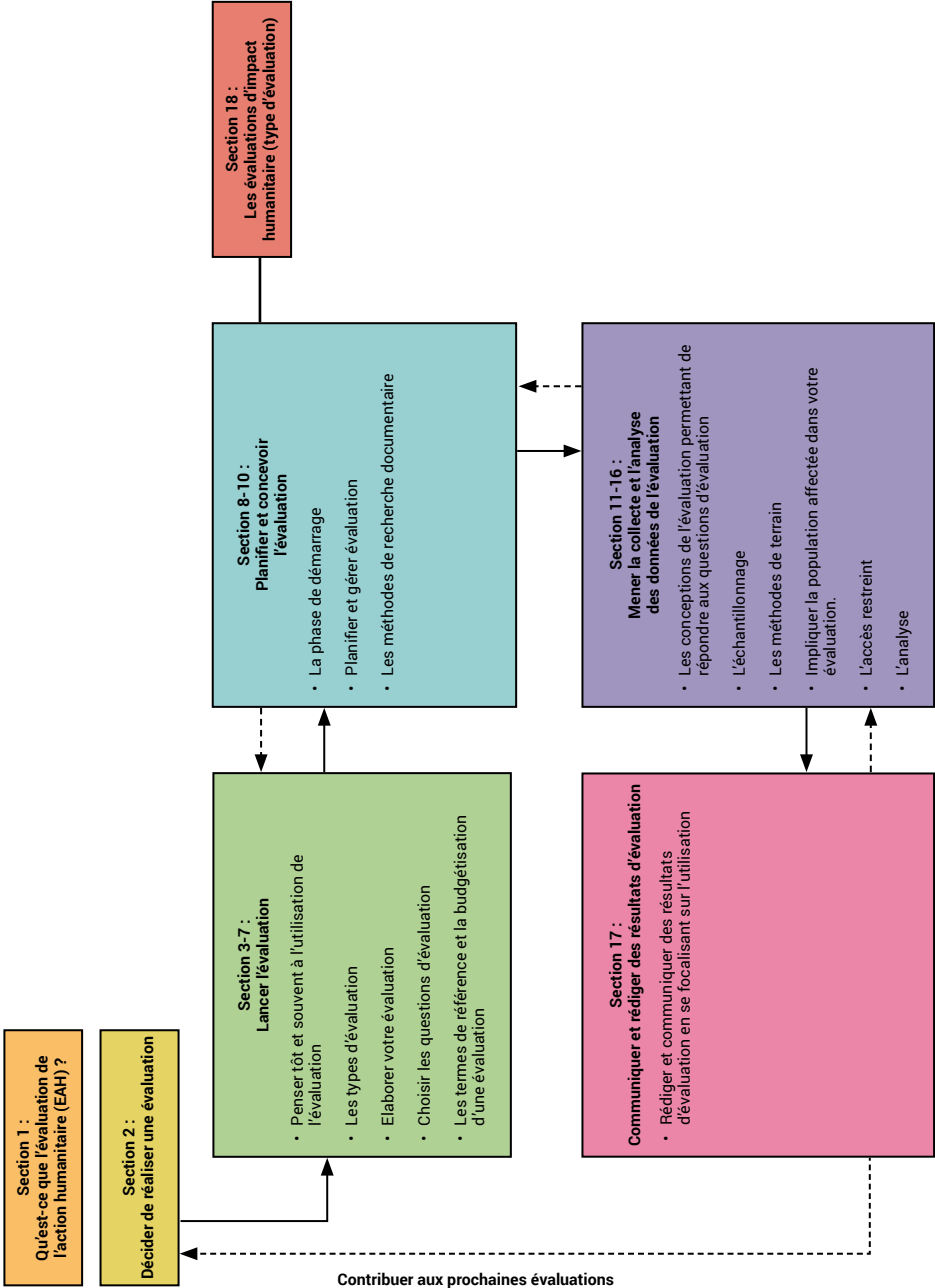
Ce Guide EAH est organisé de manière à refléter un processus d'évaluation classique. Il accompagne l'utilisateur à travers toutes les étapes que l'organisation commanditaire, le responsable de l'évaluation et l'évaluateur franchissent suite à la décision de réaliser une évaluation et jusqu'à la diffusion des résultats. Ce Guide est organisé en cinq chapitres qui reflètent les étapes importantes du processus d'évaluation puis divisé en sections qui parcourent les activités spécifiques de chacune des étapes. Ces étapes et sections sont indiquées dans le plan du Guide EAH (voir page ci-contre). Ce plan peut servir d'outil de navigation pour le Guide, mais il présente également le processus d'évaluation dans son ensemble, y compris avec les boucles de rétroaction.

Les processus d'évaluation de qualité ne sont pas linéaires. Tout d'abord, ce Guide EAH souligne l'importance de se concentrer sur l'utilisation. Les résultats d'une évaluation doivent alimenter le travail humanitaire à venir, en nourrissant les prises de décision, en facilitant l'apprentissage et en identifiant des domaines à améliorer, notamment en indiquant d'autres possibles recherches ou questions d'évaluation. Pour refléter cela, les étapes de l'évaluation sont représentées sous la forme d'une boucle ouverte plutôt que d'une ligne du temps.

Par ailleurs, les processus d'évaluation comprennent un certain nombre de boucles de retours, et des allers-retours peuvent se produire entre deux activités. Au sein de ce Guide, ces liens entre sections sont mis en exergue par des références croisées, comme celui-ci.

La bibliographie et le sommaire sont disponibles à l'adresse suivante : [www.alnap.org/EHA](http://www.alnap.org/EHA).

# Plan du Guide EAH : Comment naviguer à l'intérieur des phases et des sections de l'EAH



## Légende des éléments graphiques

### Icônes



Questions-clés



Définition



Pour approfondir



Exemple de bonnes pratiques



A garder à l'esprit



Astuce

### Tableaux et encadrés



Les tableaux et encadrés de la couleur de la section doivent être lus comme du contenu du Guide

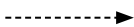


Les tableaux et encadrés en gris sont des exemples de pratiques qui viennent appuyer le texte

### Graphiques



Direct



Indirect

# Qu'est-ce que l'évaluation de l'action humanitaire ?



# 1 / Qu'est-ce que l'évaluation de l'action humanitaire ?

## 1.1 Crises humanitaires et action humanitaire

Déclenchées par des causes nombreuses et différentes, les crises humanitaires peuvent présenter des caractéristiques très distinctes. Néanmoins, les quatre manières les plus classiques de les classer et de les décrire sont les suivantes :

1. **Cause** : par exemple, un conflit armé ; une catastrophe naturelle à déclenchement rapide comme un séisme ou une inondation ; une catastrophe naturelle à déclenchement lent comme une sécheresse ; une crise sanitaire comme Ebola (FICR, s.d.).
2. **Echelle de temps et fréquence** : par exemple, prolongée, récurrente, à déclenchement rapide ou lent (ALNAP, 2007).
3. **Zone affectée** : par exemple, urbaine, rurale, mégapole (FICR, 2010).
4. **Capacité étatique/nationale pour répondre** : par exemple, si les capacités de l'Etat sont limitées, les organisations internationales peuvent jouer un rôle de remplacement, mais là où les capacités de l'Etat sont fortes, ces organisations peuvent adopter un rôle plus collaboratif (Ramalingam et Mitchell, 2014).

Il existe de nombreuses définitions de l'**action humanitaire** en réponse à une crise. L'une des plus complètes et des plus largement utilisées est celle de la Good Humanitarian Donorship Initiative (2003).



### Définition : Action humanitaire

Les objectifs de l'action humanitaire sont de sauver des vies, d'atténuer les souffrances et de préserver la dignité humaine pendant et après des crises provoquées par l'homme ou des catastrophes naturelles, ainsi que de prévenir de telles situations et d'améliorer la préparation à leur survenue.

Les paramètres de ce que l'on appelle action humanitaire se sont progressivement élargis. Autrefois, le terme désignait uniquement le fait de sauver des vies mais aujourd'hui, l'importance de sauver des moyens de subsistance est aussi largement reconnue. Ainsi, l'action humanitaire comprend à la fois l'assistance et la protection. Bien que la protection des citoyens relève clairement du rôle de l'Etat, il est également attendu que l'ensemble de la communauté humanitaire joue un rôle - et pas seulement les agences spécialisées des Nations unies et du Mouvement



international de la Croix-Rouge et du Croissant-Rouge. Notre définition de travail de l'action humanitaire renvoie également au fait de préserver la dignité humaine, l'autonomie étant une partie intégrante de cette dignité.



**Définition : Protection**

« La protection comprend l'ensemble des activités destinées à faire pleinement respecter les droits des personnes, conformément à la lettre et à l'esprit des branches pertinentes du droit » (IASC, 2011).

L'action humanitaire implique de répondre à une crise, de soutenir la préparation et la réduction des risques de catastrophes (RRC) avant une crise, ainsi que le relèvement et la réhabilitation après la crise – même si la préparation et le relèvement se situent entre les activités humanitaires et les activités de développement à long terme. L'importance de traiter les besoins de relèvement immédiatement après une catastrophe naturelle est de plus en plus reconnue. Dans le cadre des conflits et des crises prolongées, la limite entre la fin de l'urgence et le début du relèvement est souvent floue. Dans la pratique, ces deux types de soutien sont souvent nécessaires et fournis de façon simultanée.



**Astuce :**

Réfléchissez attentivement au périmètre de votre évaluation. Sur quoi voulez-vous vous concentrer ? Etes-vous plus intéressé par la première phase de la réponse d'urgence ou par une phase ultérieure ? Souhaitez-vous également observer la préparation ? Ainsi que le soutien au relèvement et à la réhabilitation ? Plus vous ciblez votre évaluation et plus vos conclusions seront susceptibles d'être utilisables.

L'action humanitaire doit être guidée par les principes d'humanité, d'impartialité, de neutralité et d'indépendance (voir [Tableau 1](#)) qui visent à différencier l'action humanitaire d'autres activités, notamment celles menées par des acteurs politiques et militaires. Ces principes sont importants pour que l'action humanitaire soit acceptée sur le terrain par les acteurs pertinents – par exemple, de manière à sécuriser l'accès aux personnes affectées par la crise (OCHA, 2012). De nombreuses organisations humanitaires ont adhéré à ces principes, souvent en exprimant leur engagement vis-à-vis du Code de conduite pour le Mouvement international de la Croix-Rouge et du Croissant-Rouge et pour les ONG lors des opérations de secours en cas de catastrophe (FICR CICR, 1994). Les principes humanitaires constituent par conséquent un point de référence clé pour évaluer l'action humanitaire. Ils doivent également guider la manière dont l'évaluation est réalisée (UNEG HEIG, 2016).

Tableau 1 : Les principes de l'action humanitaire

Humanité	Neutralité	Impartialité	Indépendance
La souffrance humaine doit être traitée partout où elle existe. L'objectif de l'action humanitaire est de protéger la vie et la santé, ainsi que de garantir le respect des êtres humains.	Les acteurs humanitaires ne doivent pas prendre parti dans les hostilités ou s'engager dans des polémiques de nature politique, raciale, religieuse ou idéologique.	L'action humanitaire doit être menée en se fondant uniquement sur les besoins, en donnant la priorité aux cas de détresse les plus urgents, sans faire de distinctions fondées sur la nationalité, la race, le genre, les croyances religieuses, la classe sociale ou les opinions politiques.	L'action humanitaire doit être autonome par rapport aux objectifs politiques, économiques, militaires ou autres que tout acteur peut avoir dans des domaines où l'action humanitaire est en cours de mise en œuvre.

Source : OCHA (2012: 2)

Les organisations humanitaires suivent également le principe du *Do No Harm* (« ne pas nuire »). Dans la Charte humanitaire, il est formulé dans le Principe de protection 1 : « éviter d'exposer à d'autres préjudices, par vos activités, la population touchée par une catastrophe », ce qui comprend non seulement les actes de violence et les violations des droits, mais aussi les dangers physiques (Sphère, 2011). Ces éléments sont davantage explicités dans la Section 2 : Décider de réaliser une évaluation. Dans la pratique courante, le *Do no harm* a été utilisé pour signifier le fait d'éviter ou de réduire tous les effets négatifs d'une intervention sur la population affectée – par exemple, installer une latrine trop près d'un puits (Bonino, 2016). Comme développé dans la Section 2 : Décider de réaliser une évaluation, ce principe doit également être appliqué à la manière dont une évaluation de l'action humanitaire (EAH) est conduite.

## 1.2 Evaluation de l'action humanitaire

Le Comité d'aide au développement de l'Organisation de coopération et de développement économique (OCDE/CAD) définit l'évaluation comme : « l'examen systématique et objectif d'un projet, d'un programme ou d'une politique en cours ou terminé(e), ainsi que leur conception, leur mise en œuvre et leurs résultats... afin de déterminer la pertinence et l'accomplissement des objectifs, l'efficience en matière de développement, l'efficacité, l'impact et la durabilité. Une évaluation devrait fournir des informations crédibles et utiles permettant d'intégrer les leçons de l'expérience dans le processus de décision des bénéficiaires et des bailleurs de fonds. L'évaluation se réfère aussi au procédé de détermination de la valeur ou de la signification d'une activité, d'une politique ou d'un programme. » (OCDE-CAD, 2002)

Partant de ce constat, nous définissons l'EAH comme :



**Définition : Evaluation de l'action humanitaire (EAH)**

L'examen systématique et objectif de l'action humanitaire afin de déterminer la valeur ou la signification d'une activité, d'une politique ou d'un programme, cela dans le but de tirer des leçons permettant d'améliorer la politique et la pratique, ainsi que d'augmenter la redevabilité ».

Une étude plus approfondie des termes clés de cette définition révèle les éléments suivants :

- **Examen** – recherche ou analyse en vue de déterminer la valeur ou la signification de l'action.
- **Systématique** – une approche planifiée et cohérente qui s'appuie sur des méthodes crédibles.
- **Objectif** – prendre du recul par rapport à l'immédiateté de l'action humanitaire et y réfléchir, en fondant les résultats sur des données probantes crédibles.
- **Tirer des leçons** pour améliorer les politiques et la pratique, mais aussi augmenter la redevabilité, est la raison d'être d'une évaluation

Les deux objectifs clés de l'évaluation sont donc les suivants : l'apprentissage et la redevabilité.



**Définition : Apprentissage**

Processus par lequel l'expérience et la réflexion mènent à des changements de comportement ou à l'acquisition de nouvelles capacités.



**Définition : Redevabilité**

La redevabilité est le moyen par lequel le pouvoir est utilisé de façon responsable. Processus qui consiste à se considérer comme responsable d'une intervention et à se montrer redevable envers les différentes parties prenantes, y compris les personnes affectées par l'intervention, celles qui la financent et les autres acteurs humanitaires.

La mesure dans laquelle une évaluation est réellement indépendante dépend de son objectif (voir Section 2 : Décider de réaliser une évaluation). Ce point est particulièrement crucial pour les évaluations tournées vers la redevabilité. Cela peut être plus difficile de mettre en œuvre pour les évaluations tournées vers l'apprentissage si les personnes qui s'occupent de l'apprentissage sont impliquées dans l'évaluation et ont été responsables de l'action humanitaire en cours d'évaluation. Même dans de tels cas, il est souhaitable d'inclure un certain niveau d'objectivité dans le processus, par exemple en ayant recours à un facilitateur externe ou à des personnes ressources expérimentées et indépendantes comme membre ou chef de l'équipe.

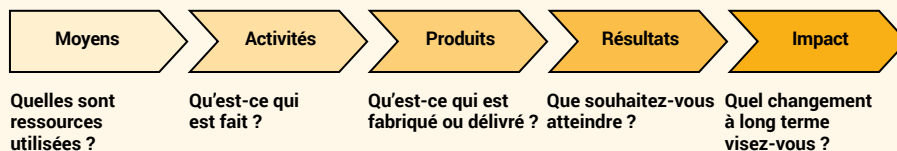


### A garder à l'esprit

Toutes les évaluations doivent avoir pour but de réduire les éléments qui peuvent la rendre biaisée.

## Les concepts clés de l'évaluation

Schéma 1.1 : La chaîne de résultats



Source : Norad (1999)



### Définition : Moyens

Les ressources financières, humaines et matérielles utilisées dans l'action humanitaire.



### Définition : Produits

Constructions, biens et services qui résultent d'une intervention.

Les produits sont le résultat de moyens reçus et d'activités conduites par l'acteur ou un groupe d'acteurs. Un produit doit être pleinement attribuable à un acteur ou un groupe d'acteurs – par exemple, les points d'eau fournis par une organisation dans un camp de personnes déplacées internes (PDI).



### Définition : Résultats

Changements ou transformations intentionnel(le)s ou non intentionnel(le)s de la situation causé(e)s directement ou indirectement par une intervention. Ceux-ci peuvent être souhaités (positifs) ou indésirables (négatifs). Ils englobent les changements de comportement (actions, relations, politiques, pratiques) des individus, groupes, communautés, organisations, institutions ou autres acteurs sociaux.

Un résultat n'est que partiellement attribuable à l'acteur responsable de l'intervention - par exemple, l'usage fait de l'eau provenant de points d'eau nouvellement installés (consommation domestique, consommation animale ou autres activités liées aux moyens de subsistance comme la fabrication de briques).





### Définition : Impact

S'intéresse aux effets plus larges du programme – sociaux, économiques, techniques et environnementaux – sur les individus, les groupes de genre et d'âge, les communautés et les institutions. Les impacts peuvent être intentionnels et non intentionnels, positifs et négatifs, macro (sectoriels) et micro (ménage, individu), à court ou long terme.

Les impacts peuvent être simultanément positifs et négatifs. Par exemple, la mise en œuvre d'une aide alimentaire peut éviter que des ménages ne vendent leurs moyens de production, comme leur bétail, pour acheter de la nourriture (impact positif intentionnel), mais cela peut également gêner la production alimentaire locale (impact non intentionnel et potentiellement négatif).

**Note: Tout au long de ce Guide, le terme « impact » est utilisé pour décrire les effets les plus larges de l'action humanitaire, comme décrit dans le Schéma 18.1 de la page 357 (Smutylo, 2001), et conformément à la définition de l'impact de l'OCDE-CAD.**



### Définition : Attribution

L'imputation d'un lien de causalité entre des changements observés (ou que l'on s'attend à observer) et une intervention spécifique.

Lors d'interventions humanitaires complexes, il est rarement possible d'attribuer un résultat à une seule cause. Une organisation spécialisée dans l'aide alimentaire peut attribuer une réduction de la malnutrition à des distributions alimentaires mais cette réduction peut aussi être due à une amélioration de la qualité de l'eau, des pratiques de soins infantiles, de l'hygiène, des soins de santé, de l'assainissement, du contrôle des vecteurs ou même des variations saisonnières normales. Comme le démontre le Schéma 18.1 de la page 357, l'attribution se complique au fur et à mesure que l'on avance sur la chaîne des résultats – il est ainsi plus difficile d'attribuer des impacts à une intervention spécifique que d'attribuer des résultats.



### Définition : Contribution

Analyser la contribution dans le cadre de l'évaluation revient à montrer de façon crédible qu'une intervention a joué un certain rôle dans l'obtention de résultats. L'analyse de la contribution est une sorte d'analyse évaluative qui reconnaît que plusieurs causes peuvent contribuer à un résultat, même si individuellement elles peuvent ne pas être nécessaires ou suffisantes pour avoir un impact.

Dans le domaine de l'évaluation de l'action humanitaire, il est en général beaucoup plus facile d'évaluer la contribution que l'attribution.

## 1.3 Suivi et évaluation dans le cycle de réponse humanitaire

L'évaluation des projets et des programmes humanitaires est habituellement une activité unique réalisée à un moment clé du cycle de la réponse humanitaire d'urgence dans le but de fournir des informations à ce cycle ainsi que de nourrir les réponses futures.<sup>1</sup> Dans certains cas, une série d'évaluations peut être planifiée à différentes étapes du cycle de la réponse, comme pour l'évaluation du Darfour (Broughton *et al.*, 2006) et la réponse au séisme en Haïti (Grünewald *et al.*, 2010 ; Hidalgo et Theodate, 2012). Par ailleurs, un suivi doit être réalisé tout au long de la mise en œuvre des projets et des programmes humanitaires. Le suivi et l'évaluation (S&E) sont des outils complémentaires pour aider à déterminer dans quelle mesure une intervention fonctionne bien (FICR, 2010:19). Comme l'explique l'Agence suédoise de coopération internationale au développement (SIDA) : « pour qu'une évaluation soit faisable, il se peut toutefois que des données de suivi soient nécessaires. Si une intervention n'a pas été suivie depuis le départ de façon adéquate, il sera peut-être impossible de réaliser ensuite une évaluation satisfaisante. Tout comme le suivi a besoin de l'évaluation comme complément, l'évaluation a besoin du soutien du suivi » (Molund et Schill, 2007: 15)



### Définition : Suivi

Processus continu recourant à la collecte systématique de données à partir d'indicateurs spécifiés pour fournir des indications sur l'étendue des progrès, l'atteinte des objectifs et le niveau de consommation des fonds alloués aux gestionnaires et aux principales parties prenantes d'une intervention humanitaire en cours. (d'après OCDE-CAD, 2002)

En règle générale, les personnes qui mettent en œuvre un programme sont responsables de son suivi pour s'assurer qu'il reste sur la bonne voie.<sup>2</sup> Par exemple, le suivi d'un programme d'aide alimentaire d'urgence se concentrera probablement sur les moyens tels que les quantités d'aide alimentaire distribuée, et sur les produits, comme le nombre de personnes recevant l'aide alimentaire. Il s'intéresse aussi parfois aux résultats et à l'impact – par exemple, si les individus vendent la nourriture (c'est-à-dire comment ils l'utilisent – le résultat) et si cela a un impact au niveau des prix des denrées alimentaires sur le marché. Dans un programme de transfert monétaire, le suivi peut s'intéresser à la façon dont les individus ont reçu l'argent aussi bien qu'au moment où cela a eu lieu et au montant reçu. Il se focalise alors sur les moyens et les produits. Le suivi peut également s'étendre aux résultats – par exemple, en s'intéressant à ce que les individus ont fait de l'argent reçu.

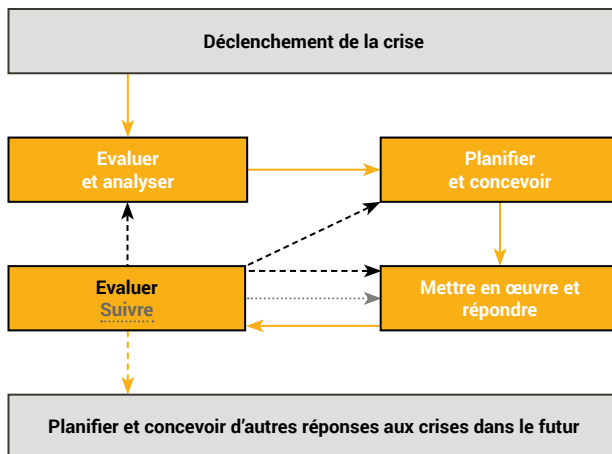
Le suivi est souvent faible dans l'action humanitaire et le manque de données de suivi fiables crée régulièrement des problèmes lors de l'évaluation. Une évaluation est en général conduite par des personnes qui ne sont pas impliquées dans la mise en œuvre du programme afin de lui conférer une meilleure crédibilité.

L'évaluation se focalise particulièrement sur les résultats et l'impact : dans l'exemple du programme de transfert monétaire évoqué ci-dessus, elle peut observer les conséquences du transfert à destination des femmes, ainsi que des impacts plus larges comme l'impact sur les activités des marchés et du commerce.

Le [Schéma 1.2](#) montre comment le S&E se situe à l'intérieur du cycle de réponse d'urgence. Comme le soulignent Hallam et Bonino (2013), les défis auxquels est confrontée l'EAH sont le reflet des défis et des complexités que rencontre l'action humanitaire. Certains des défis les plus courants auxquels est confrontée l'EAH sont décrits dans la [Section 1.4 : Les défis les plus courants](#).

Cependant, toutes les évaluations de l'action humanitaire ne se rapportent pas aux programmes et au cycle de la réponse d'urgence. Elles peuvent également se focaliser sur d'autres éléments comme la politique ([Section 4 : Les types d'évaluation](#)). Comme développé dans la [Section 3 : Penser tôt et souvent à l'utilisation de l'évaluation](#), il est important de prendre en compte le lieu où viennent se placer les résultats d'une évaluation au sein d'un programme plus large ou des processus d'une organisation. Pour garantir qu'une évaluation est utile et utilisée, il faut planifier le moment où ses résultats pourront le mieux contribuer à des instants clés de la prise de décision.

Schéma 1.2 : Suivi et évaluation dans le cycle de réponse d'urgence



## 1.4 Les défis les plus courants de l'EAH

L'EAH fait face à deux grands types de défis : ceux communs à toutes les évaluations, souvent renforcés dans des contextes humanitaires, et ceux spécifiquement associés à l'évaluation de l'action humanitaire, qui est menée en général dans des environnements difficiles. Ce *Guide* vise à traiter les défis propres à l'EAH. La liste qui suit décrit certains des défis les plus courants et leurs solutions éventuelles, en relation avec la section où ils sont présentés dans ce *Guide*.<sup>3</sup>

Les évaluateurs doivent clarifier les contraintes qu'ils rencontrent dès le début de l'évaluation et la manière dont ils essaient de les traiter du point de vue de la gestion des risques. Le rapport final doit indiquer comment ces contraintes ont affecté le processus et les conclusions de l'évaluation.

### L'urgence et le chaos des secours humanitaires

Une réponse humanitaire rapide est, par définition, planifiée rapidement et souvent in extremis. Les documents de planification et de suivi peuvent être rares, les objectifs imprécis, et les plans de départ devenir rapidement obsolètes à mesure que le contexte change ou est clarifié. Par conséquent, quel(le) référence ou point de départ allez-vous utiliser pour votre évaluation ?

#### **Solution potentielle :**

pensez à construire une chronologie de la crise et de l'action humanitaire à évaluer. Utilisez des entretiens avec le personnel de l'organisation pour identifier des objectifs réels ou implicites et la manière dont ils auraient pu changer au fil du temps. Voir [Section 10 : Les méthodes de recherche documentaire](#).

### L'insécurité peut être synonyme d'un manque d'accès dans les environnements de conflit

Des questions de sécurité peuvent compliquer ou rendre impossible, en particulier dans des environnements de conflit, l'atteinte de la population affectée par les évaluateurs. Comment pouvez-vous mener une évaluation sans accès à la population affectée ?

#### **Solution potentielle :**

Explorez des façons innovantes de réaliser l'évaluation à distance. Soyez clair sur toutes les contraintes que vous avez rencontrées lorsque vous présentez vos conclusions et rédigez votre rapport. Assurez-vous de ne pas généraliser vos conclusions à propos de lieux et de populations que vous n'avez pas pu atteindre. Voir [Section 15 : L'accès restreint](#).



## Manque de données de référence

Les données peuvent avoir été détruites lors de la crise, ou ne plus être pertinentes (par exemple, quand une grande proportion de la population a été déplacée). A l'intérieur du cours laps de temps d'une évaluation, comment collectez-vous toutes les données dont vous avez besoin, sans valeur de référence ?

### Solution potentielle :

Réalisez des entretiens pour demander aux personnes affectées par la crise et aux informateurs locaux dans quelle mesure les conditions ont changé et pourquoi, utilisez par exemple le rappel des faits. Incluez des questions similaires dans les enquêtes. Voir [Section 14 : Impliquer la population affectée dans votre évaluation et Encadré des bonnes pratiques page 270](#).

## Fort renouvellement du personnel

Le fort renouvellement du personnel qui travaille dans l'action humanitaire, en particulier le personnel expatrié, complique le travail des évaluateurs pour ce qui est de trouver et d'interroger les informateurs clés. Comment trouvez-vous le personnel clé pour la période que vous évaluez ?

### Solution potentielle :

Invitez les informateurs clés qui faisaient partie du personnel – ou sont absents pendant la période de l'évaluation - à participer à un entretien téléphonique (ou Skype) ou à une enquête en ligne. Appuyez-vous davantage sur le personnel national qui présente en général des taux de renouvellement moindres, ce qui peut impliquer de traduire les enquêtes en ligne, etc. Voir [Section 13 : Les méthodes de terrain](#).

## Crise humanitaire située dans des zones isolées comprenant des infrastructures endommagées

Certaines crises humanitaires se produisent dans des zones isolées et/ou là où des infrastructures ont été endommagées, ce qui peut rendre l'accès difficile. Comment allez-vous atteindre ces zones et comment planifiez-vous le travail de terrain ?

### Solution potentielle :

Planifiez avec soin le travail de terrain avec une personne qui connaît bien la situation actuelle, y compris les probables temps de transport et contraintes d'accès. Ou alors, pensez à confier à l'organisation partenaire le calendrier des visites de terrain en fonction des critères définis par l'équipe d'évaluation. Prévoyez des moments réservés aux imprévus dans votre plan. Voir [Section 15 sur l'accès restreint](#).

## Les conflits polarisent les points de vue

Les conflits renforcent souvent les différences de points de vue. Les événements, à commencer par la crise elle-même, peuvent être sujets à différentes interprétations. Comment pouvez-vous réaliser des évaluations « objectives » dans de tels contextes ?

### Solution potentielle :

Familiarisez-vous avec les lignes de fracture du conflit et rassemblez autant de points de vue différents que possible, en particulier de groupes situés dans des camps opposés. Assurez-vous que ceux-ci sont représentés dans le rapport d'évaluation final. Il peut être difficile de garantir l'indépendance de l'évaluation dans ces circonstances. Si vous pensez que l'évaluation est compromise – par exemple, vous n'avez pas obtenu l'accès à certaines zones contestées –, assurez-vous de clarifier ce point dans le rapport d'évaluation et de prendre en compte ses implications dans vos conclusions. Voir [Section 14 : Impliquer la population affectée dans votre évaluation](#).

## Perte de confiance en raison de la politisation ou d'un traumatisme

Lors de conflits et d'autres crises humanitaires qui se sont politisées et où l'on retrouve de manière généralisée des maltraitances, des traumatismes et de la peur, la confiance peut être perdue au sein de la population affectée. Comment allez-vous obtenir des informations approfondies et précises de la population affectée durant le court laps de temps que dure une évaluation ?

### Solution potentielle :

Concevez des méthodes de collecte de données et des manières d'impliquer la population affectée qui soient sensibles aux traumatismes et aux peurs. Pensez à choisir l'intervieweur en fonction du sexe de la personne interrogée. Voir [Section 11 : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation](#) et [Section 13 : Les méthodes de terrain](#).

## Les acteurs humanitaires travaillent dans des environnements sous pression et stressants

Les organisations humanitaires et leur personnel peuvent rencontrer des difficultés pour mettre en œuvre leurs programmes dans des environnements à fort stress. Prendre du temps pour le consacrer à des évaluateurs peut alors être vu comme non prioritaire, en particulier lors d'[évaluations en temps réel](#) au début de la crise. Quelles méthodes sont alors appropriées dans un tel contexte ?

### Solution potentielle :

Soyez sensible aux pressions auxquelles les acteurs de l'aide sont confrontés et trouvez des façons « légères » de les impliquer. Par exemple, de courts ateliers de réflexion et d'apprentissage peuvent apporter de bons éclairages s'ils sont bien animés. Voir [Section 2 : Décider de réaliser une évaluation](#). Demandez au personnel ce qu'il veut savoir à propos de son intervention et dans quels domaines il est désireux de changer la manière dont son organisation aborde une tâche.

### Pression horaire sur la population affectée

Les personnes peuvent avoir peu de temps à consacrer à une évaluation parce que leur propre survie est leur priorité. Comment pouvez-vous garantir que la population affectée participe à l'évaluation de façon appropriée ?

#### Solution potentielle :

Évitez les méthodes chronophages comme les *focus groups*, et optez plutôt pour la consultation de membres de la population affectée lorsqu'ils attendent des distributions ou un transport. De plus, utilisez l'observation directe. Prenez en compte ces pressions horaires au moment de la conception de l'évaluation : par exemple, est-il possible de reporter l'évaluation après le pic de la réponse de manière à ce que la population soit plus apaisée et subisse moins de pression ? Voir [Section 14 : Impliquer la population affectée dans votre évaluation](#).

### Manque de responsabilités clairement définies entre acteurs humanitaires

Le manque de responsabilités clairement définies au sein des acteurs humanitaires peut gêner la redevabilité et l'attribution de l'impact, en particulier lors d'une crise de grande échelle impliquant de nombreux acteurs humanitaires. Comment allez-vous attribuer les résultats et l'impact ?

#### Solution potentielle :

Concentrez-vous sur la contribution plutôt que sur l'attribution. Pensez à réaliser une évaluation conjointe pour explorer l'impact combiné d'un certain nombre d'organisations, ou de la réponse humanitaire internationale dans son ensemble, plutôt que d'essayer d'isoler artificiellement l'impact d'une seule organisation. Voir [Section 4 : Les types d'évaluation](#).


### Termes de référence (TdR) trop ambitieux et ressources limitées

Les TdR peuvent impliquer une charge de travail irréaliste, par exemple s'il est attendu qu'une tâche demande l'équivalent de deux personnes travaillant sur une période d'un an alors que le budget équivaut à une personne pendant quatre mois (voir [Section 7](#) qui traite des Termes de référence). Comment allez-vous gérer les attentes et réaliser une évaluation crédible ?

#### Solution potentielle :

Utilisez un rapport de démarrage pour préciser le travail d'évaluation de manière à ce qu'il puisse être terminé dans le temps et le budget alloués. C'est une manière utile de gérer les attentes en amont. Voir [Section 8 : La phase de démarrage](#).

### Contraintes éthiques concernant les approches expérimentales

L'hypothèse selon laquelle « si les personnes affectées par la sécheresse n'avaient pas reçu d'aide alimentaire, beaucoup d'entre elles seraient mortes » est souvent appelée contrefactuelle. Des considérations éthiques interdiraient bien sûr de tester cette hypothèse qui consisterait à priver certaines personnes de l'aide. La Section 2 et la Section 14 traitent les considérations éthiques de l'EAH.) D'autres considérations sur cette matière sont présentées tout au long de ce Guide comme .

#### Solution potentielle :

Comparez les différentes méthodes d'aide (par exemple, les dons en espèces, les bons d'achats, la distribution alimentaire). Profitez de toutes les « expériences naturelles » qui ont pu se produire, comme celle d'un groupe qui n'a pas reçu d'aide en raison de son isolement. Voir [Section 11 : Les conceptions de l'évaluation](#) permettant de répondre aux questions d'évaluation.

### Evaluer la protection

Les défis renvoient ici au manque de clarté quant à la définition de la protection, ainsi qu'à la difficulté de mesurer le non-quantifiable – en particulier ce qui ne s'est pas produit, par exemple des actes de violence ou des violations de droits, comme résultat d'actions menées par des organisations humanitaires. **Voir Companion Protection Guide** (Christoplos et Bonino, 2016)

Malgré ces défis considérables, l'EAH doit souvent être entreprise rapidement. Elle peut par conséquent nécessiter des compétences évaluatives supérieures à celles d'autres évaluations (voir Section 9 : Planifier et gérer votre évaluation), de manière à garantir que la rigueur et la crédibilité ne sont pas compromises (UNICEF, 2013: 3-4).

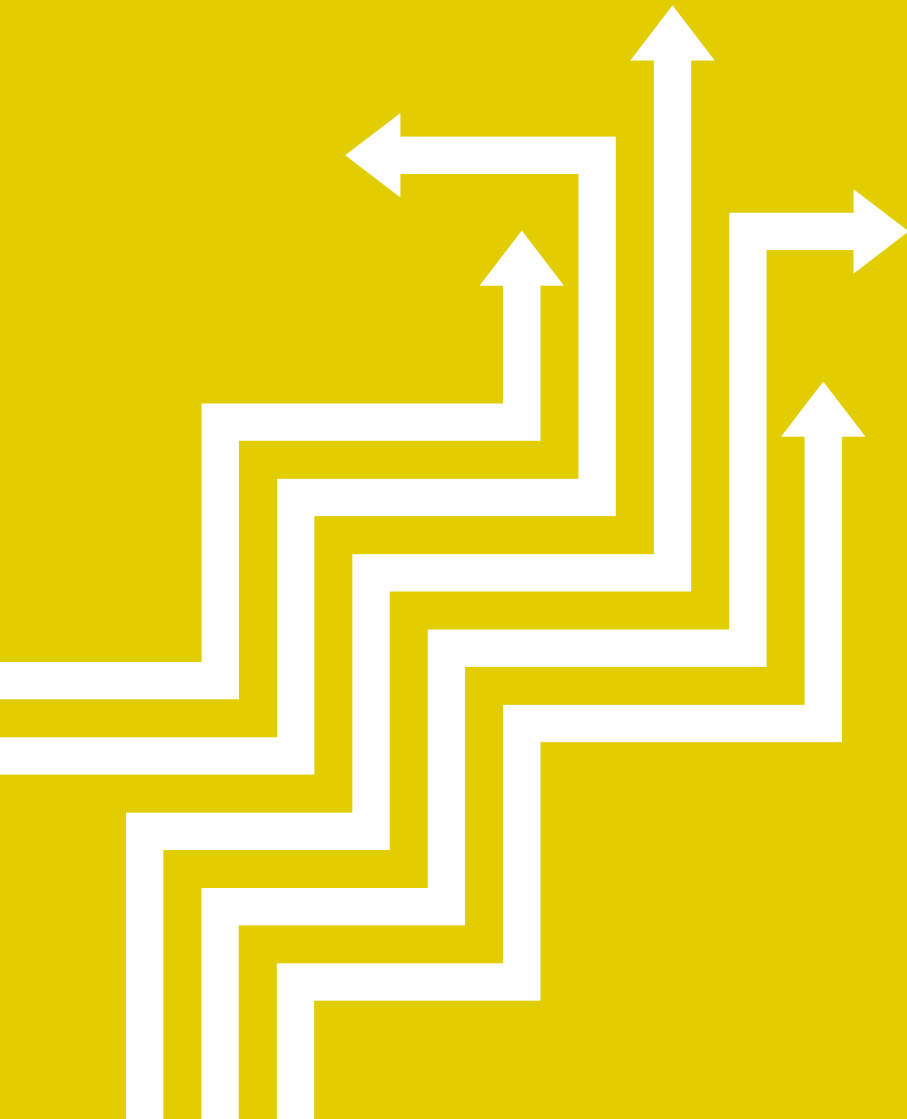
## Notes de fin de texte

1. Cette section s'appuie sur Morra Imas et Rist (2009) ainsi qu'Hallam (2011).
2. Il existe des formes extérieures de suivi, comme les rapports d'évaluation citoyenne (Citizen Report Cards de la Banque mondiale, 2004), mais elles sont généralement utilisées dans les contextes de développement plutôt qu'humanitaires. Le suivi par une tierce partie devient plus courant dans les urgences complexes comme en Syrie (voir le travail de SAVE à ce sujet : <http://www.save.gppi.net/home/>).
3. Section adaptée de Beck (2006).

## Notes

## Notes

# Décider de réaliser une évaluation





# 2 / Décider de réaliser une évaluation

Commanditer une évaluation est parfois une obligation contractuelle envers un bailleur, mais dans d'autres cas l'évaluation peut être mise dans la balance avec d'autres options. Cette section présente quelques options s'appuyant sur l'objectif de cette activité évaluative. Si vous êtes un responsable de programme ou un responsable d'évaluation, cette section vous aidera à décider si une évaluation est bien la meilleure option, et si oui, quand la réaliser. Elle vous fournira également des conseils pour savoir comment mener une « étude de faisabilité de l'évaluation ».

## 2.1 Quand l'évaluation convient-elle à des objectifs de redevabilité et d'apprentissage ?

### Redevabilité

L'évaluation est l'un des nombreux processus permettant à une organisation de s'acquitter de sa redevabilité, autrement dit de sa responsabilité vis-à-vis des autres, que ce soit son conseil d'administration, ses bailleurs ou les populations affectées. Vous devez décider si c'est la solution la plus appropriée et la plus rentable.

Le [Tableau 2.1](#) résume les différents types de redevabilité et suggère d'autres solutions en plus de l'évaluation. Par exemple, si la redevabilité et la conformité financières sont les principales préoccupations – des bailleurs, de l'équipe de direction ou du conseil d'administration de l'organisation -, un audit peut être plus approprié. Un audit examine en effet la conformité aux exigences et est ainsi différent d'une évaluation qui délivre un jugement de valeur (FICR, 2010). La redevabilité envers les populations affectées (RPA) fait désormais l'objet d'une attention beaucoup plus grande, comme expliqué dans la [Section 14 : Impliquer la population affectée dans votre évaluation](#). Cette redevabilité est principalement obtenue grâce à un processus continu de retours et de consultation durant la mise en œuvre (Darcy *et al.*, 2013). Les évaluations peuvent examiner la RPA et doivent également chercher à soutenir ce type de redevabilité (voir HAP, 2010), mais il est trop tard pour la prendre en compte lorsque l'EAH a débuté.

La redevabilité présente souvent des aspects à la fois externes et internes. La direction d'une organisation peut être redevable de l'utilisation des ressources envers les bailleurs (externes) et envers son conseil d'administration (internes) puisqu'une mauvaise utilisation des ressources peut menacer la survie de l'organisation. De la même manière, la redevabilité envers la population affectée peut refléter des obligations que la direction et le conseil d'administration se sont engagés à respecter.

D'après Sandison (2006) :

L'impact de l'évaluation est renforcé, voire rendu possible, s'il fait partie d'une plus large gamme d'approches visant à améliorer la performance. Le suivi, par exemple, reste le parent pauvre de l'évaluation alors qu'il doit recevoir la même attention de la part des décideurs. L'évaluation, quelle que soit sa forme, n'est qu'un élément de la redevabilité. (p. 140)



**Questions-clés : Si la redevabilité est le principal objectif de votre évaluation, demandez-vous :**

- Quel type de redevabilité vous intéresse le plus, et envers qui (référence au [Tableau 2.1](#)) ?
- Une évaluation est-elle le meilleur moyen de procéder ?
- Une évaluation doit-elle être liée à d'autres processus de redevabilité ? (Voir [exemple de bonnes pratiques ci-dessous](#))



**Exemple de bonnes pratiques : Lier les mécanismes de redevabilité : audit et évaluation**

Pour deux de ses évaluations pays, l'organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) a réalisé un audit peu de temps avant l'évaluation indépendante (Somalie en 2012 et Soudan du Sud en 2015). Cet enchaînement a bien fonctionné dans le sens où les questions de rentabilité et de conformité, en particulier en lien avec l'efficacité et l'efficience des processus et des procédures de gestion, ont été couvertes par l'audit avant l'évaluation. Cela a permis à l'équipe d'évaluation de se focaliser sur des questions plus stratégiques de programmation, tout en s'appuyant sur certains résultats de l'audit.

Tableau 2.1 : L'évaluation et les autres processus de redevabilité

Type de redevabilité	L'évaluation est-elle appropriée ?	Autres processus de redevabilité à prendre en compte
Redevabilité envers la population affectée	Oui, si l'évaluation est conçue de façon appropriée, avec une pleine implication de la population affectée.	Le dialogue et la consultation continus avec la population affectée, ainsi que l'utilisation de mécanismes de retours tout au long de la vie du projet/ programme, peuvent être des manières plus efficaces d'être redevables envers la population affectée que le recours à une évaluation.
Redevabilité stratégique (par exemple envers le mandat et les objectifs de l'organisation)	Oui	Analyse stratégique
Redevabilité managériale (par exemple, pour l'utilisation des ressources au sein d'une organisation)	Oui	Gestion de la performance et autres outils de gestion Audit de performance
Redevabilité (par exemple, envers les bailleurs) et conformité (par exemple, envers la direction et le conseil d'administration) financières	Oui, en particulier si la rentabilité et l'efficacité sont les principaux sujets d'intérêt	Audit (peut être plus approprié si le contrôle et la conformité des finances sont les principaux sujets d'intérêt)
Redevabilité contractuelle (par exemple, envers ses engagements)	Oui, en particulier s'il existe une obligation contractuelle à faire une évaluation.	Audit et autres processus spécifiés dans le contrat
Redevabilité relationnelle (par exemple, envers d'autres organisations impliquées dans une opération)	Oui, si incluse dans les <u>Termes de référence</u> de l'évaluation	Analyse institutionnelle
Redevabilité juridique (par exemple, envers les droits locaux ou internationaux)	Non	Examen du respect des conditions légales (par exemple, code du travail)

Dans la pratique, les organisations de bailleurs sont parmi les parties prenantes les plus puissantes. De ce fait, la « redevabilité ascendante » envers les bailleurs tend à motiver une part importante de l'activité évaluative (Brown et Donini, 2014). Les évaluations tournées vers la redevabilité peuvent également être utilisées pour faire des rapports à la direction et au conseil d'administration. L'accent est mis sur les résultats, observant souvent dans quelle mesure les ressources ont été correctement utilisées pour remplir le mandat et les responsabilités d'une organisation. Pour cette raison, les évaluations tournées vers la redevabilité ont en général lieu à mi-parcours ou vers la fin d'un programme, moment à partir duquel des résultats sont disponibles et peuvent être analysés dans un rapport.

### Apprentissage

Les évaluations tournées vers l'apprentissage sont conçues pour faciliter l'apprentissage individuel, de groupe et/ou organisationnel. Elles constituent d'excellentes opportunités pour un apprentissage qui s'appuie sur des cas réels et peuvent être très efficaces en ce qui concerne l'analyse de ce qui a fonctionné ou non, et la manière dont les performances peuvent être améliorées. Les évaluations tournées vers l'apprentissage se focalisent souvent sur des programmes qui en sont à leur phase initiale et de mise en œuvre afin d'identifier ce qui fonctionne bien et ce qui peut être amélioré. Elles peuvent avoir lieu à tout moment du cycle du programme. Dans le secteur humanitaire en particulier où les ressources pour l'apprentissage sont rares et où l'on capitalise peu sur l'apprentissage (ALNAP, OCDE-CAD Evaluation Network et UNEG, 2010 ; Patrick, 2011 ; Knox-Clarke et Darcy, 2014), les évaluations peuvent être très utiles pour créer de la connaissance et lancer des processus d'apprentissage organisationnel.

L'évaluation n'est pourtant pas le seul processus d'apprentissage ni le plus rentable. D'autres processus à prendre en compte sont présentés dans le [Tableau 2.2](#). Certains peuvent être intégrés dans une évaluation tournée vers l'apprentissage – par exemple, les « revues après action » (Voir [Section : 13.7](#))  
Les méthodes tournées vers l'apprentissage).

Tableau 2.2 : Les processus d'apprentissage

Méthode d'apprentissage	Breve description	Comment cette méthode peut être utilisée dans le cadre d'une évaluation
<u>Revue après action</u>	Processus facilité par un animateur pour que les personnes impliquées dans le programme puissent réfléchir sur ce qui s'est passé, les réussites, les difficultés et les leçons à en tirer.	La revue après action peut être facilitée par les évaluateurs, en tant que partie de l'évaluation tournée vers l'apprentissage. L'apprentissage peut être intégré dans le rapport d'évaluation (USAID, 2006 ; Ramalingam, 2006: 64).
<u>Le storytelling qui utilise les métaphores</u>	Processus facilité où les participants « racontent l'histoire » de ce qui s'est passé et en tirent les leçons.	Le <i>storytelling</i> qui utilise les métaphores peut être utilisé avec le personnel, comme décrit dans l'exemple de bonnes pratiques page 260, de manière à encourager la réflexion et l'apprentissage partagés. Better Evaluation propose un certain nombre de références utiles (Better Evaluation, 2014).
<u>La technique du « changement le plus important »</u>	Dans le cadre de cette approche participative, les histoires de « changement important » sont collectées au niveau du terrain. Les plus significatives sont sélectionnées par des groupes de parties prenantes désignées et ce, comme une façon de saisir les résultats et l'impact du projet (Davies et Dart, 2005).	La technique du « changement le plus important » peut être utilisée avec les personnes affectées par une crise de manière à identifier les changements qu'elles attribuent à l'intervention (comme utilisé par l'évaluation de la réponse de la FICR au séisme qui s'est produit au Pérou en 2007, Martinez, 2009).
Enquête d'encouragement	Approche qui se focalise sur les solutions plutôt que sur les problèmes, impliquant les parties prenantes dans un exercice de réflexion sur ce qui a fonctionné et sur les enseignements qui peuvent être tirés des réussites.	Rarement utilisée dans l'EAH mais Preskill et Catsambas (2006) analysent en détail l'enquête d'encouragement et donnent des exemples de son utilisation dans des évaluations de différents secteurs.

Certaines organisations, comme UNICEF et Tearfund, utilisent des processus d'apprentissage autonomes dans des réponses d'urgence de petite ou moyenne échelle (par exemple, les revues après action) et essaient d'encourager les ateliers de réflexion et d'apprentissage. Cependant, pour les réponses d'urgence de plus grande échelle, l'UNICEF a recours à l'évaluation, notamment aux ETR (voir [Section 4 : Les types d'évaluation](#)).

Une évaluation seule n'aboutira pas à de l'apprentissage organisationnel, mais les personnes impliquées dans l'évaluation – et à ses côtés – peuvent apprendre en tant qu'individus. Une *organisation apprenante* ([Tableau 2.3](#)) est plus susceptible de tirer profit d'une évaluation. Le travail d'ALNAP intitulé *L'utilisation pour un vrai changement* (Hallam et Bonino, 2013) identifie le soutien de la direction, une culture

organisationnelle favorable à l'évaluation et des structures organisationnelles qui la promeuvent, ainsi que la capacité à garantir des ressources humaines et financières adéquates, comme fondamentales pour les organisations humanitaires qui utilisent véritablement les évaluations.

**Tableau 2.3** : L'évaluation et les trois piliers d'une organisation apprenante

Pilier	Caractéristiques distinctives	Implications pour l'évaluation
Un environnement propice à l'apprentissage	Les membres de l'équipe se sentent libres d'être en désaccord entre eux, de poser des questions candides, d'admettre leurs erreurs et de représenter les points de vue minoritaires. Ils reconnaissent la valeur des idées opposées et sont prêts à prendre des risques ainsi qu'à explorer des domaines inconnus.	Utiliser les méthodes d'évaluation qui encouragent l'ouverture d'esprit comme l'enquête appréciative, la technique dite du « changement le plus important » et les revues après action.
Processus d'apprentissage concrets	Des processus formels existent pour générer, collecter, interpréter et diffuser de l'information.	Demander éventuellement au personnel de faire partie de l'équipe d'évaluation. Définir un processus formel pour impliquer le personnel dans l'analyse des conclusions et les leçons tirées.
Une direction qui renforce l'apprentissage	La haute direction de l'organisation démontre sa volonté d'alimenter des points de vue alternatifs ; pointe l'importance de consacrer du temps à l'identification du problème, au transfert de connaissances et à la réflexion ; s'engage dans un questionnement et une écoute actifs.	Permet que l'évaluation soit fortement soutenue au niveau de l'encadrement et bénéficie d'une durée adéquate. Crée un espace pour les retours et la discussion des résultats avec la direction générale.

Source : Adapté de Garvin *et al.* (2008)

## 2.2 L'évaluation est-elle le bon outil pour le travail demandé ?

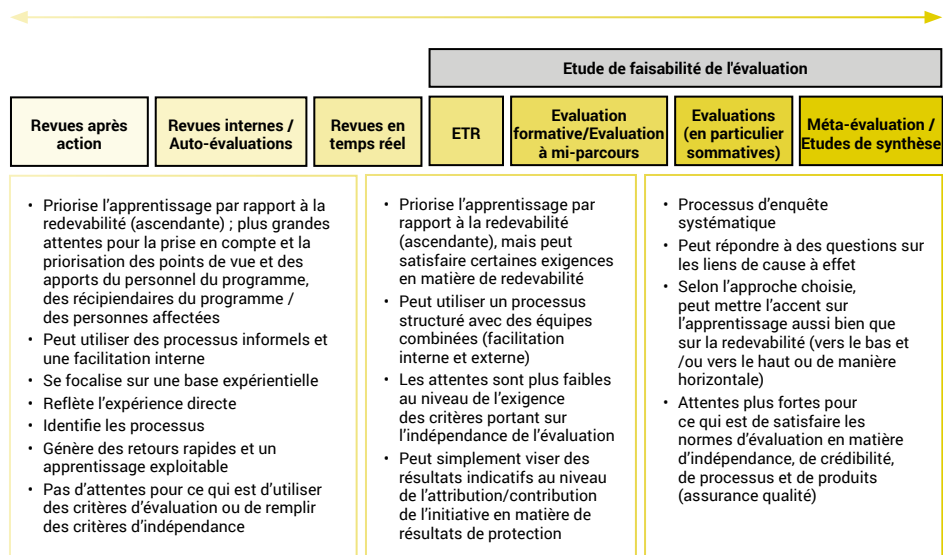
Etes-vous certain que vous voulez réaliser une évaluation ? L'évaluation coûte de l'argent qui aurait pu être utilisé pour une action humanitaire permettant de sauver des vies ou d'atténuer des souffrances. L'argent de l'évaluation est bien dépensé s'il permet d'améliorer l'action humanitaire, mais il peut le faire seulement si les conclusions de l'évaluation sont suivies d'effets. Il est par conséquent important de choisir le bon outil pour le travail demandé.

Le Schéma 2.1 illustre l'éventail des options depuis les exercices informels de réflexion et d'apprentissage (sur la gauche) jusqu'aux approches d'évaluation structurées, nécessaires en particulier pour les évaluations tournées vers la redevabilité (sur la droite). Lorsque vous avez clarifié l'objectif général de l'évaluation, il se peut que vous souhaitiez réfléchir à la position de votre activité évaluative.

Schéma 2.1 Eventail des options depuis les exercices de réflexion et d'apprentissage jusqu'aux approches évaluatives plus formelles

**Permettent une structuration moindre et plus d'informalité :** moins d'exigences strictes pour respecter les normes d'évaluation et utiliser les critères d'évaluation

**Demandent une structuration plus importante :** Attentes en matière de respect des normes au niveau du processus de conception et d'analyse ainsi que pour les réalisations attendues



Source : Bonino (2016), développé à partir de Scharbatke-Church (2011: 7)

## 2.3 Equilibrer apprentissage et redevabilité

Comme le suggère le [Schéma 2.1](#), de nombreuses évaluations servent en pratique le double objectif de redevabilité et d'apprentissage. Comme l'énonce Irene Guijt : « Vous ne pouvez pas être redevable si vous n'apprenez pas. Et vous devez savoir dans quelle mesure vous êtes à la hauteur de vos attentes en matière de performance si vous voulez apprendre » (2010: 277), même s'il peut être difficile en pratique d'atteindre les deux objectifs. Néanmoins, il est important de déterminer quel est l'objectif le plus important des deux ou celui qui domine car cela aidera à déterminer le choix de la conception de l'évaluation.

Lors de la planification et de la conception de l'évaluation, voici les questions clés à se poser :



### Questions-clés

- Quel objectif est le plus important – la redevabilité ou l'apprentissage ? Où, dans le continuum entre redevabilité et apprentissage, placeriez-vous l'objectif ? Celui-ci ne doit pas être au milieu si vous ne voulez pas un résultat flou.
- Si la redevabilité est l'objectif principal : redevabilité envers qui et dans quel but ?
- Si l'apprentissage est l'objectif principal : par qui ? Et comment l'apprentissage est-il censé se dérouler ?

### Les caractéristiques des évaluations tournées vers la redevabilité et l'apprentissage

Une fois que vous avez clarifié l'objectif principal, vous pouvez concevoir le processus et les méthodes d'évaluation permettant de l'atteindre. Certaines caractéristiques correspondent à un objectif d'évaluation. Par exemple, une évaluation tournée vers la redevabilité est susceptible de s'intéresser davantage à l'objectivité et à l'indépendance, et d'adopter un style plus proche de l'enquête, cherchant à attribuer la responsabilité des réussites comme des échecs. Pour autant, cela peut ne pas favoriser l'apprentissage si les personnes qui doivent apprendre se sentent sur la défensive. L'apprentissage demande un environnement psychologique sûr (Edmondson, 2014) où l'on accepte de reconnaître les difficultés et d'admettre les erreurs. Ainsi, une évaluation orientée vers l'apprentissage utilisera probablement un style comportant plus d'animation, encourageant les personnes impliquées dans la mise en œuvre à participer et réfléchir. Ces différences expliquent pourquoi il peut être difficile de combiner apprentissage et redevabilité dans l'objectif d'une seule et même évaluation. Voir le [Tableau 2.4](#) décrivant comment l'approche et le style de l'évaluation peuvent varier entre une évaluation orientée vers la redevabilité et une autre orientée vers l'apprentissage.



Tableau 2.4 : Les évaluations orientées vers la redevabilité et l'apprentissage

Élément de l'évaluation	Évaluation tournée vers la redevabilité	Évaluation tournée vers l'apprentissage
Termes de référence	Basée sur les contributions des parties prenantes externes ainsi que des responsables du programme	Devrait être mise en place par les personnes directement impliquées dans le programme qui souhaitent ou ont besoin d'apprendre
Composition de l'équipe	En général, personnel extérieur indépendant	Personnel interne, avec éventuellement un facilitateur ou un chef d'équipe externe, ou une combinaison de personnels interne et externe
Approche privilégiée	Méthodologie de collecte et d'analyse des données (plus objective)	Processus de réflexion et d'élaboration de résultats (plus participative)
Méthodes	Mélange de méthodes quantitatives et qualitatives qui fourniront des données probantes solides	Méthodes participatives impliquant les personnes qui doivent apprendre
Direction	Les personnes en charge de la redevabilité	Les personnes responsables de la gestion des connaissances et de l'apprentissage
Mode de direction	Plus directif	Plus axé sur l'animation
Circulation du rapport	En général, dans le domaine public	Peut être limitée à l'organisation pour encourager une participation ouverte et franche

## 2.4 Décider quoi et quand évaluer

Pour s'assurer qu'une évaluation est utile et utilisée, elle doit être planifiée au moment où ses résultats peuvent au mieux contribuer à des instants clés de prise de décision. Comme l'expliquent Hallam et Bonino (2013), le *timing* de l'évaluation en tant que tel et de la présentation de ses conclusions influence considérablement la demande d'évaluation. Il est donc important que les commanditaires d'évaluations prennent ce point en compte. De nombreux utilisateurs potentiels se plaignent qu'elles arrivent souvent trop tard pour être d'une quelconque utilité dans la prise de décision (Weiss, 1983b ; 1990).

Ainsi, un rapport sur les travaux de recherche et les évaluations du DFID constate : Au sein des personnes interrogées, la critique la plus courante des évaluations était le timing : même si elles constituent des moments importants pour les bureaux nationaux, on considère qu'elles prennent en général trop de temps pour être utiles aux équipes en charge des politiques, et que l'on accorde une attention insuffisante au fait de les relier aux cycles des politiques et aux fenêtres d'opportunités. (Jones et Mendizabal, 2010: 11)

Il est contreproductif de commander des évaluations pour chaque projet ou programme si elles sont susceptibles de surcharger l'organisation et si des systèmes de suivi efficaces sont en place. Des évaluations choisies de façon stratégique sont plus à même de satisfaire des besoins clés en matière de connaissance dans un délai convenable.

Les questions clés à prendre en compte pour des évaluations choisies de façon stratégique sont :



#### Questions-clés

- Comment le processus d'évaluation peut-il présenter une valeur ajoutée pour l'organisation dans son ensemble ?
- Combien d'évaluations l'organisation a-t-elle la capacité d'absorber ?

L'unité d'évaluation de l'Agence suédoise de coopération internationale au développement (SIDA) a expérimenté avec succès l'adoption d'une approche stratégique pour les évaluations humanitaires et de développement. Cette approche pourrait de la même manière s'appliquer uniquement aux EAH.



#### Exemple de bonnes pratiques : Sélection stratégique des évaluations

Le cycle de planification des évaluations de l'Agence suédoise de coopération internationale au développement commence avec les échanges que l'unité d'évaluation mène auprès de toutes les unités opérationnelles afin de déterminer ce qu'elles aimeraient savoir et comment une évaluation pourrait les aider en cela. Une liste d'environ 100 idées d'évaluations est tout d'abord compilée, liste à partir de laquelle le département de l'évaluation en choisit 15. Celles-ci sont à ajouter aux 80 évaluations réalisées chaque année au niveau opérationnel (SIDA, 2011).

Le Norwegian Refugee Council (NRC) priorise les évaluations « sur la base de la pertinence – évaluer les bons éléments au bon moment », en les identifiant au moyen d'un plan d'évaluation annuel (NRC, 2015).

Il peut être particulièrement approprié de mener une évaluation :

- Pour un programme dont les résultats ne sont pas connus ou sujets à débats
- Pour des interventions de grande taille et coûteuses
- Pour des initiatives pilotes, ou pour tester une nouvelle idée de programme
- Là où l'organisation a des intérêts stratégiques
- Là où les parties prenantes souhaitent une évaluation

Il est inapproprié de réaliser une évaluation :

- Quand il est peu probable qu'elle apporte de nouvelles informations
- Quand des questions de sécurité ou de manque de données pourraient réduire sa crédibilité

Weiss (1998) ajoute certaines mises en garde importantes quant au moment où il peut être futile de mener une évaluation à cause des dynamiques organisationnelles internes. Parmi les exemples, on trouve celui où le responsable de l'évaluation (ou son équivalent) limite de façon stricte ce que l'évaluation peut étudier, excluant certains sujets importants, ou celui où l'évaluation ne dispose pas de ressources financières suffisantes ou humaines suffisamment qualifiées pour être réalisée dans de bonnes conditions.

Planifier tôt l'évaluation dans la mise en œuvre d'un programme humanitaire signifie que des systèmes de suivi pertinents peuvent être mis en place dès le début. Cela peut également renseigner et influencer la conception de l'évaluation finale.

Certaines organisations ont des mécanismes de mise en œuvre automatique ou systèmes de déclencheur pour certains types d'évaluation, en particulier les agences des Nations unies. Par exemple, le seuil de déclenchement d'une évaluation à la FAO est la dépense de 10 millions de dollars ou plus, et pour l'UNICEF les « urgences humanitaires majeures ». Ces mécanismes de mise en œuvre automatique sont principalement associés à de la redevabilité, lorsque d'importantes sommes d'argent ont été dépensées, et dans certaine mesure aux cas de potentiels risques en matière de réputation.

Beaucoup de ces aspects peuvent être saisis dans le cadre d'une politique d'évaluation. Une politique d'évaluation peut également guider de nombreuses décisions mises en exergue dans cette section portant sur l'objectif de l'évaluation, s'il faut en faire une et quand, ainsi que sur les considérations d'ordre éthique et les questions d'assurance qualité à prendre en compte. Si vous commanditez une évaluation, vous devez indiquer la politique d'évaluation de votre organisation dans les termes de référence (TdR) de l'évaluation. Les évaluateurs doivent se familiariser avec la politique d'évaluation dès le début car elle peut contenir des instructions et des exigences au niveau de la conception et des méthodes. Voir Exemple de bonnes pratiques, page 52, à propos de la politique d'évaluation succincte mais complète du NRC.



### Exemple de bonnes pratiques : La politique d'évaluation du NRC

La politique d'évaluation du NRC « clarifie pourquoi, comment et quand le NRC a recours à des évaluations. Elle précise également les normes de conduite et définit comment les évaluations contribuent à l'apprentissage et à la redevabilité de l'organisation » (p. 3). C'est un document de 11 pages destiné au personnel en charge de commanditer, planifier et utiliser les évaluations. Il expose les quatre principes qui sous-tendent son travail d'évaluation : pertinence, flexibilité, transparence et participation, en reconnaissant les défis spécifiques de l'évaluation du travail du NRC avec les PDI, qui peuvent avoir subi des événements traumatiques, ainsi que d'autres défis fréquents de l'EAH comme l'accès restreint et les contextes qui évoluent rapidement. Ce document de politique indique les exigences minimums pour les évaluations, par exemple le fait que tous les bureaux nationaux doivent commanditer au moins une évaluation externe tous les trois ans. Il fournit enfin des conseils pour s'assurer que les évaluations sont utiles et indique les normes de qualité à suivre.

Source : NRC (2015)

### Réaliser une étude de faisabilité de l'évaluation<sup>1</sup>

Il est utile de réaliser une étude de faisabilité de l'évaluation *avant* de décider de lancer une évaluation, cela afin de prendre en compte certaines des questions soulevées ci-dessus. Cette étude est un processus descriptif et analytique visant à produire une base raisonnée pour procéder à une évaluation (Schmidt, Scanlon et Bell, 1979). Elle peut également être utilisée pour proposer des solutions permettant de maximiser la faisabilité et l'utilité d'une évaluation à commanditer à un stade ultérieur (Davies, 2013). Le principal objectif de cette étude est à la fois de décider si l'évaluation peut être entreprise ou non, et de s'assurer que des mesures sont prises durant la mise en œuvre de l'intervention de manière à ce que les conditions soient en place pour faciliter le processus d'évaluation (UNEG 2011 : 17). Si l'étude conclut que l'évaluation doit avoir lieu, le travail fait dans le cadre de l'étude de faisabilité peut alimenter de façon directe les termes de référence de l'évaluation.

Même si c'est rare dans l'EHA, une étude de faisabilité de l'évaluation peut être réalisée peu après le début d'une intervention humanitaire de manière à améliorer la conception du projet et ainsi sa capacité ultérieure à être évaluée, par exemple en mettant en place des systèmes de suivi efficaces et en suivant des indicateurs définis. Dans l'EAH, les études de faisabilité d'évaluations sont plus courantes pour déterminer si le *timing* est bon pour l'évaluation, ou si elle doit être reportée à une phase ultérieure ou jusqu'à ce que le contexte soit plus opportun, par exemple lorsque la sécurité s'est améliorée.

Rick Davies (2013) a synthétisé les conseils disponibles en matière d'études de faisabilité d'évaluations de façon à produire le résumé suivant des principaux sujets couverts et des mesures habituellement suggérées (Tableau 2.5).

Tableau 2.5 : Cadre générique pour une analyse de l'évaluabilité

Principaux problèmes /étapes	Étapes le plus souvent proposés
Définir le périmètre du projet	<ul style="list-style-type: none"> <li>• Définir la période, l'extension géographique, et les acteurs concernés</li> <li>• Accord sur les résultats attendus de l'étude d'évaluabilité</li> </ul>
Identifier les ressources disponibles	<ul style="list-style-type: none"> <li>• Identifier les documents</li> <li>• Identifier les parties prenantes</li> </ul>
Identifier et examiner les documents	<ul style="list-style-type: none"> <li>• La logique/théorie du changement/chaîne de résultats du programme ainsi que la clarté, la vraisemblance et l'appropriation</li> <li>• Systèmes d'information</li> <li>• Disponibilité, pertinence et la qualité des données, capacité des systèmes et du personnel de faire le nécessaire</li> <li>• Analyser les mises en œuvre par rapport aux objectifs initiaux</li> </ul>
S'impliquer avec les parties prenantes	<ul style="list-style-type: none"> <li>• Identifier ce que les acteurs comprennent de l'objectif du programme, de sa conception et de sa mise en œuvre, en prenant en compte les points d'accord et de désaccord</li> <li>• Identifier leurs attentes concernant une évaluation : objectifs, processus et utilisation</li> <li>• Clarifier et combler les lacunes recensées dans les documents</li> </ul>
Établir des conclusions et faire des recommandations	<p>Les conclusions et les recommandations de l'analyse de l'évaluabilité devraient inclure :</p> <ul style="list-style-type: none"> <li>• Les améliorations de la logique du projet</li> <li>• Le renforcement des systèmes et des capacités de S&amp;E (si cela est demandé et comme cela a été demandé)</li> <li>• Questions de l'évaluation d'un intérêt prioritaire pour les acteurs</li> <li>• Potentielle conception de l'évaluation seulement si cela est demandé et comme cela a été demandé *</li> </ul>
Partager les constatations et conclusions avec les parties prenantes	

\* Davies met en garde à ceux qui commanditent ou mettent en œuvre des analyses de l'évaluabilité contre la tentation de s'aventurer sur les questions liées à la conception en commençant à faire des suggestions sur les potentielles conceptions.  
Source : Davies, 2013: 16

Dans le cadre de l'EAH, les études de faisabilité d'évaluations peuvent couvrir de façon utile quatre domaines clés :

1. **Niveau d'ambition général et type de questions** auquel les parties prenantes de l'évaluation et les parties prenantes du programme aimeraient que l'évaluation réponde.
2. **Conception du programme et logique d'intervention** – ce point est particulièrement important pour les évaluations de résultats et d'impacts qui utilisent des conceptions à base théorique pour comprendre la causalité, ainsi que pour des conceptions avec des méthodes combinées et des approches fondées sur les résultats qui observent la contribution dans le cadre d'interventions multi-acteurs ou en réseaux (par exemple, la cartographie des incidences ; la collecte des résultats ; RAPID Outcome Mapping Approach – ROMA).

3. **Disponibilité des données** ou faisabilité de générer des données avec les ressources allouées, de façon à ce que l'évaluation puisse répondre aux questions évaluatives retenues.
4. **Caractère favorable du contexte** pour réaliser une évaluation, par exemple en termes d'accès, de logistique et de sécurité, ainsi que la capacité du bureau local à accueillir l'équipe d'évaluation, ou encore le « climat » organisationnel et le soutien de la direction en faveur de l'évaluation.

Le quatrième domaine – le caractère favorable du contexte (voir [Section 15 : l'accès restreint](#)) – est la raison la plus à même de vous décider à ne pas lancer une évaluation ou à la reporter. Les études de faisabilité des évaluations peuvent être utiles s'il n'est pas clair qu'une évaluation crédible sera possible – cf. l'étude de faisabilité de l'évaluation en Somalie (Cosgrave, 2010). Ces études sont parfois combinées avec les examens initiaux des programmes, comme ce fut le cas avec l'examen du programme et l'étude de faisabilité de l'évaluation du programme de l'UNICEF en faveur de l'éducation dans les situations d'urgence et de transition post-crisis (Barakat *et al.*, 2010) ainsi que pour l'examen préparatoire du Pôle Mondial d'éducation (Reid *et al.*, 2010).

Dans la pratique, une étude de faisabilité de l'évaluation spécifique est en général utilisée pour des programmes humanitaires de grande échelle et des évaluations multi-agences. Pour la plupart des EAH, et en particulier pour les évaluations de petite échelle, il est en général suffisant que l'équipe d'évaluation réalise un rapide examen de « l'évaluabilité » durant la phase de démarrage. Cette étude peut recommander des améliorations au niveau du périmètre et de l'angle de l'évaluation. Une fois l'équipe d'évaluation désignée, il est en général trop tard pour révéifier le caractère favorable du contexte et pour envisager de décaler ou reporter l'évaluation. Si vous commanditez une EHA de petite échelle et avez des inquiétudes quant au caractère favorable du contexte, vous devez envisager de faire un rapide examen de « l'évaluabilité » avant de désigner les membres de l'équipe d'évaluation.



#### **Exemple de bonnes pratiques : Un examen rapide de l'évaluabilité**

Durant la phase de démarrage d'une évaluation conjointe commanditée au Népal par le réseau Communicating with Disaster Affected Communities (CDAC) des initiatives Communicating with Communities (CwC) en réponse au séisme de 2015, l'équipe d'évaluation a estimé que le but global et les objectifs définis dans les termes de référence demandaient une étude de recherche pour tester les hypothèses plutôt qu'une évaluation. Cela a été discuté avec le Groupe de pilotage de l'évaluation conjointe durant la phase de démarrage, et un consensus s'est établi quant au fait qu'il fallait bien envisager une étude de recherche et non une évaluation.<sup>2</sup>

Une étude ou un rapide examen de la faisabilité de l'évaluation peuvent ainsi garantir que l'EAH est adaptée au contexte – par exemple, le type de crise et le type de réponse à évaluer, ainsi que le contexte organisationnel, notamment dans quelle mesure il existe une culture de l'apprentissage et un engagement de la direction.

## 2.5 Ethique et EAH

Diverses structures d'évaluation ont produit des guides d'éthique en la matière. A titre d'exemple, les principes directeurs à destination des évaluateurs développés par l'Association américaine de l'évaluation insistent sur le questionnement permanent, la compétence, l'intégrité et l'honnêteté, le respect des personnes et la responsabilité vis-à-vis du bien-être public et général.<sup>3</sup> Les lignes directrices éthiques du Groupe d'évaluation des Nations unies (2008) se focalisent quant à elles sur l'intentionnalité de l'évaluation, les obligations des évaluateurs, les obligations envers les participants, le processus et le produit de l'évaluation.

Certaines organisations humanitaires ont développé des lignes directrices éthiques spécifiquement pour l'EAH. Par exemple, la Fédération internationale de la Croix-Rouge et du Croissant-Rouge (FICR) se réfère au principe du « *Do no harm* », qui est davantage explicité [page 56](#), et renvoie également au principe 5 du *Code de conduite pour le Mouvement international de la Croix-Rouge et du Croissant-Rouge et pour les ONG lors des opérations de secours en cas de catastrophe* : « nous nous appliquerons à respecter la culture, les structures et les coutumes des communautés et des pays dans lesquels nous menons nos activités » (FICR, 2010).

La politique éthique du NRC implique de :

- Ne pas nuire aux informateurs et aux parties prenantes ;
- Suivre les normes internationales lorsque l'on s'entretient avec des enfants et des personnes ayant subi des violences à caractère sexiste, et
- Se focaliser sur les questions évaluatives et éviter de perturber les personnes (par exemple, en ne leur demandant pas de revivre des événements traumatiques). (NRC, 2015)

Prendre en compte ces lignes directrices d'ordre éthique peut être important lorsque l'on réalise une étude de faisabilité de l'évaluation. A titre d'exemple, le travail de terrain peut-il être mené de manière à ce que les informateurs et les parties prenantes ne soient pas mis en danger ? Ce sujet est abordé [dans l'encadré page 57](#). L'éthique liée à l'implication de la population affectée dans l'EAH est davantage développée dans la [Section 14 : Impliquer la population affectée dans votre évaluation](#).

## L'évaluation doit respecter le principe du « Do no harm » (ne pas nuire)

Un certain nombre de considérations d'ordre éthique existent dans l'EAH, mais elles doivent toujours suivre le principe du « Do no harm », tout comme devraient le faire les opérations humanitaires qui sont évaluées. Cela s'avère particulièrement pertinent pour les évaluations réalisées dans des situations de conflit et de forte insécurité. Le point de départ doit être le fait d'envisager comment le processus de l'évaluation peut affecter les personnes qui y prennent part ou sont consultées.<sup>4</sup>

Qu'une évaluation étudie une réponse à un conflit ou une catastrophe naturelle, les évaluateurs doivent être conscients de la manière dont elle peut exacerber des tensions en :

- Créant des attentes par rapport au fait que prendre part à l'évaluation (par exemple, durant la collecte des données) aboutira à l'arrivée de plus d'aide, ce qui peut générer de la frustration ;
- Déclenchant des discussions tendues, par exemple entre différents groupes au sein de la population affectée, durant le processus de collecte des données, en soulevant des questions qui renforcent les tensions et les divisions ;
- Étant perçus comme impliqués dans la collecte de renseignements par l'une des parties au conflit ;
- Présentant de façon involontaire une analyse biaisée sur la base d'une représentation inadéquate des points de vue des différentes parties prenantes.

Pour éviter de tels risques et réaliser l'évaluation d'une manière qui soit « sensible au conflit », les responsables de l'évaluation et les évaluateurs doivent depuis le début envisager de :

- Étudier si toute étape du processus d'évaluation peut contribuer à des tensions (cela nécessitera de se concentrer en particulier sur la collecte des données et la diffusion du rapport)
- Pour les situations de conflit, réaliser une nouvelle analyse du conflit (ou une mise à jour de l'existante) de manière à alimenter la planification et la conception de l'EAH (voir Chigas et Goldwyn, 2012)
- Passer en revue les plans d'évaluation à la lumière de cette analyse de manière à s'assurer qu'ils ne contribuent pas à des tensions (et là où cela est possible, et si cela entre dans le mandat de l'organisation, essayer de les minimiser).

Source : Adapté de Christoplos et Bonino (2016)





## Pour approfondir : Qu'est-ce que le « *Do no harm* » (ne pas nuire) ?

La signification du terme *Do no harm* diffère selon les domaines de l'action humanitaire et la sensibilité du conflit. Le *Do no harm* est dérivé du principe médical selon lequel un praticien médical ne doit causer aucun préjudice au patient. Il a été adopté et adapté dans d'autres domaines.

**Du point de vue humanitaire**, le *Do no harm* est un terme communément utilisé mais souvent mal utilisé. Dans le *Manuel Sphère*, il est évoqué dans le Principe de protection 1 : « éviter d'exposer à d'autres préjudices, par vos activités, la population touchée par une catastrophe », ce qui comprend les actes de violence et les violations des droits, mais aussi les dangers physiques. Dans la pratique courante, le *Do no harm* a parfois été utilisé pour signifier le fait d'éviter ou de minimiser tout effet négatif d'une intervention, par exemple l'installation d'une latrine trop près d'un puits a souvent été décrite comme un « préjudice ».

**Du point de vue de la sensibilité au conflit**, le *Do no harm* est un cadre spécifique en sept étapes qui peut être utilisé pour évaluer la sensibilité au conflit d'une intervention. Il a été développé par Collaborative for Development Action (aujourd'hui CDA), et s'avère « l'outil » le plus largement utilisé pour évaluer la sensibilité au conflit.

La sensibilité au conflit signifie s'assurer qu'une intervention ne contribue pas de façon involontaire au conflit et, là où cela est possible, contribue à la paix (dans le cadre du mandat d'une organisation). Dans cette définition, le *Do no harm* est associé aux risques liés à un conflit, notamment les nombreux risques liés à la protection, puisque ceux-ci sont étroitement imbriqués.

Il est important de noter que la boîte à outils de la sensibilité au conflit contient de nombreux autres outils au-delà du *Do no harm*, et l'on trouve beaucoup de documents pratiques et analytiques dans le domaine de la sensibilité au conflit de façon plus générale.

Source : Christoplos et Bonino, 2016

## Notes de fin de texte

1. Largement inspiré de Christoplos et Bonino (2016).
2. Voir le site Internet du CDAC Network : [www.cdacnetwork.org](http://www.cdacnetwork.org).
3. Voir [www.eval.org/p/cm/ld/fid=51](http://www.eval.org/p/cm/ld/fid=51).
4. Ces points sont également couverts dans le guide UNEG sur les obligations éthiques destinées à ceux qui initient, gèrent et réalisent des évaluations (UNEG, 2008). Ces obligations comprennent : le respect de la dignité et de la diversité ; les droits de l'Homme ; la confidentialité ; et la prévention des dommages.

## Notes

## Notes

# Lancer l'évaluation



# 3 / Penser tôt et souvent à l'utilisation de l'évaluation

Cette section vise à aider les commanditaires et les responsables d'évaluations à prendre en compte dès le début ce qu'ils doivent faire pour s'assurer que les conclusions et les recommandations des évaluations seront utilisées. Elle s'appuie fortement sur le travail d'ALNAP en matière d'utilisation de l'évaluation, lequel cherche à améliorer la manière dont les organisations humanitaires utilisent et adoptent les conclusions des EAH (Sandison 2006 ; Hallam, 2011 ; Hallam et Bonino, 2013).

## 3.1 Ce que signifie se focaliser sur l'utilisation

Une évaluation focalisée sur l'utilisation est réalisée en ayant à l'esprit les principaux utilisateurs visés et dans le but d'usages pratiques, spécifiques et déclarés.



Sandison, en s'appuyant les travaux de Patton (1997) et Carlsson *et al.* (1999), propose la définition suivante de l'utilisation : « une évaluation a été utilisée si des utilisateurs ayant l'intention et le potentiel pour agir ont pris en compte de façon sérieuse et active ses conclusions, identifiant des usages pertinents par rapport à leurs propres intérêts et besoins » (2006: 100-101).

Il est essentiel d'identifier très tôt les utilisateurs visés afin de les aider à décider ce qu'ils veulent obtenir avec l'évaluation. Leur implication dans la définition de l'objet, des objectifs et du périmètre de l'EHA doit orienter le choix de la conception et des méthodes. La focalisation sur l'utilisation et la collaboration avec les principaux utilisateurs doivent ensuite se poursuivre de manière à orienter le processus d'évaluation de la planification jusqu'à la mise en œuvre. Cette façon de penser et de travailler demande de l'engagement et du temps, mais garantit que l'EHA contribue à renforcer les connaissances des utilisateurs et à apporter des changements et des améliorations dans la pratique (Hallam et Bonino, 2013).



### A garder à l'esprit

Il est important de se souvenir que l'évaluation coûte de l'argent, argent qui aurait pu être utilisé pour une action humanitaire, permettant de sauver des vies ou d'atténuer des souffrances. L'argent de l'évaluation est bien dépensé s'il permet d'améliorer l'action humanitaire, mais il peut le faire seulement si les conclusions de l'évaluation sont suivies d'effets.

## 3.2 Comment sont utilisées les évaluations de l'action humanitaire

Les EAH sont généralement utilisées de deux façons : avec des fins de redevabilité (également appelées évaluations sommatives), et avec des fins d'apprentissage, auquel cas les évaluations se focalisent le plus souvent sur un projet spécifique (évaluations formatives) mais aussi parfois sur de la connaissance systémique et générale (évaluations évolutives). La plupart des EAH comprennent une combinaison de ces éléments, même si l'un d'eux peut prévaloir (voir [Tableau 3.1](#)).

Tableau 3.1 : Les objectifs de l'évaluation

Utilisation	Questions	Exemples
<b>Sommative</b> – juger le mérite ou la valeur d'un programme (par exemple, pour respecter sa redevabilité vis-à-vis des acteurs ou les informer des décisions de financement)	Le programme répond-il aux besoins ? A-t-il du mérite ? Quels sont ses résultats ?	L'évaluation conjointe des bailleurs de l'aide humanitaire et de reconstruction en Afghanistan entre 2001 et 2005 (DANIDA, 2005).
<b>Formative</b> – pour renforcer l'apprentissage (par exemple, améliorer un programme)	Qu'est-ce qui fonctionne et qu'est-ce qui ne fonctionne pas ? Quelles sont les forces et les faiblesses actuelles ?	The Organisational Learning Review of Caritas Internationalis' Response to the Tsunami Emergency (Otto <i>et al.</i> , 2006) a facilité le processus d'apprentissage, en soulignant l'ouverture et la participation des parties prenantes clés.  Les évaluations en temps réel ont généralement un objectif principal d'apprentissage (Saavedra, 2013).
<b>Évolutive</b> – pour contribuer à de nouveaux concepts, idées et manières de penser au sein d'une organisation ou d'un secteur dans son ensemble	Le programme prend-il en compte l'actualité et les contraintes ? Quels sont les modèles généraux que l'on retrouve parmi tous les programmes ?	L'évaluation conjointe menée au Rwanda (Borton <i>et al.</i> , 1996) a introduit de nouvelles idées concernant la redevabilité des organisations humanitaires, et a précipité l'arrivée de nombreuses innovations majeures en matière de politiques, comme les standards Sphère.

On peut classer la manière dont les EAH sont utilisées selon les quatre catégories suivantes :

- **Utilisation instrumentale** : mise en œuvre directe des conclusions et des recommandations par les décideurs, par exemple pour procéder à des changements dans un programme en cours, ou pour s'attaquer à certaines pratiques. La plupart des EAH sont conçues pour ce type d'utilisation, qu'elles soient orientées vers la redevabilité ou l'apprentissage.
- **Utilisation conceptuelle** : les résultats et les conclusions de l'évaluation sont assimilés par l'organisation sous la forme de nouvelles idées et de nouveaux concepts, comme décrit dans la catégorie « évolutive » du [Tableau 3.1](#). Cette utilisation est moins courante et se produit souvent de façon progressive. Les changements conceptuels peuvent ne pas être attribuables à une seule évaluation, mais les synthèses d'évaluation peuvent être particulièrement utiles pour déclencher un changement conceptuel.
- **Processus ou apprentissage** : l'engagement et la participation dans l'évaluation en tant que tel peuvent générer de l'apprentissage individuel et déclencher des changements au niveau du comportement.
- **Utilisation de légitimation** : l'évaluation légitime une décision ou une vision existante au sein d'une organisation ou d'un bureau national en fournissant des données probantes indépendantes et objectives susceptibles d'être utilisées pour justifier des actions ultérieures (voir Sandison, 2006 ; Hallam et Bonino, 2013: 18).

Parfois, les évaluations ne sont pas utilisées de façon efficace pour diverses raisons :

- **Utilisation ritualiste** : l'évaluation n'a de fonction que symbolique, par exemple pour remplir une obligation contractuelle vis-à-vis du bailleur, mais avec peu ou pas du tout d'engagement de la part de l'organisation pour ce qui est de l'utiliser.
- **Utilisation impropre** : des conclusions sont supprimées, mal interprétées, ou faussées afin de servir des intérêts personnels ou politiques.
- **Non-utilisation** : les conclusions sont ignorées parce que les utilisateurs leur trouvent une valeur faible ou inexistante (en raison, par exemple, de recommandations mal formulées), ou alors parce qu'ils ne les ont pas lues, ou encore parce que le contexte a radicalement changé. La non-utilisation peut être une conséquence d'un manque d'engagement de la part de la direction, d'une mauvaise conception de l'évaluation, ou d'une incapacité de l'évaluation à répondre à ses propres questions ou à fournir des preuves irréfutables de ses conclusions (Sandison, 2006).



### 3.3 Qui sont les utilisateurs visés de l'évaluation ? Quels sont leurs besoins en information et comment peuvent-ils être impliqués ?

Cette sous-section fournit des astuces pour les personnes qui commanditent et gèrent une évaluation, au niveau des phases de conception et de planification, afin de garantir qu'elle soit focalisée sur l'utilisation.

On compte ainsi trois étapes-clés :

1. Identifier les parties prenantes de l'évaluation, et au sein de ce groupe les principaux utilisateurs visés.
2. Consulter les utilisateurs visés à propos de leurs besoins d'informations tout en faisant en sorte que le périmètre de l'évaluation demeure réaliste et gérable.
3. Trouver des façons d'impliquer les utilisateurs visés tout au long de l'évaluation.

Ce processus peut également être utile pour l'équipe d'évaluation. Si ces étapes n'ont pas été respectées avant que l'équipe d'évaluation ne soit désignée, alors l'équipe doit les faciliter avant de lancer le processus d'évaluation. A titre d'exemple, durant la phase de démarrage, l'équipe d'évaluation peut animer un exercice de cartographie des parties prenantes avec le responsable de l'évaluation (voir p. 66) et consulter les principaux utilisateurs visés pour savoir ce qu'elles attendent de l'évaluation.

#### Identifier les parties prenantes

La première étape est d'identifier les parties prenantes et les utilisateurs de l'évaluation. Une partie prenante est un individu ou une organisation qui a un intérêt ou un enjeu dans l'évaluation parce qu'elle peut être affectée par des décisions issues de ses conclusions ou de ses recommandations.

Cependant, toutes les parties prenantes n'utiliseront pas nécessairement l'évaluation finale. La population affectée en est un bon exemple. Elle est une partie prenante parce que l'évaluation est une opportunité qui lui permet d'exprimer ses points de vue, ce qui peut aboutir à des améliorations du programme, mais il est peu probable que la population utilisera ses conclusions finales.

#### Approcher vos utilisateurs visés

Les utilisateurs visés sont le public principal de l'EAH. Les quatre questions suivantes vous aideront à identifier les principaux utilisateurs visés. Ce processus peut également prendre la forme d'un exercice participatif visuel ou d'une analyse des parties prenantes selon le pouvoir et l'intérêt.

## 1. Quelles sont les parties prenantes de votre évaluation?

Rappelez-vous qu'il s'agit des parties prenantes de l'évaluation, et non de l'action humanitaire en cours d'évaluation.

Les parties prenantes peuvent aller de l'unité d'évaluation qui commande l'évaluation jusqu'au personnel qui met en œuvre le programme, en passant par les autorités locales et la population affectée.

## 2. Quelles parties prenantes ont un intérêt direct dans l'évaluation ?

Les parties prenantes ayant un intérêt direct doivent être pleinement impliquées dans l'évaluation. Ces parties prenantes peuvent comprendre les financeurs et le personnel qui met en œuvre le programme.

## 2. Les parties prenantes ayant un intérêt indirect ?

Les parties prenantes ayant un intérêt indirect comprennent celles qui doivent être influencées par l'évaluation ou sont consultées (par exemple, le personnel d'autres organisations qui travaillent dans la zone), et celles qui devraient être consultées, par exemple la population affectée.

## 3. Parmi les parties prenantes ayant un intérêt direct, lesquelles sont vos utilisateurs visés ?

Dans la plupart des cas, si l'on attend qu'il apprenne des conclusions de l'évaluation, l'utilisateur visé est le personnel qui met en œuvre le programme. Les financeurs peuvent également être les utilisateurs visés dans le cadre d'une évaluation tournée vers la redevabilité.

## 4. Qui sont vos principaux utilisateurs visés ?

Cette question est particulièrement importante si l'évaluation présente plusieurs parties prenantes majeures ayant des intérêts différents et concurrents. A moins que vous n'identifiez clairement les utilisateurs visés et ne les priorisiez, les objectifs concurrents peuvent devenir ingérables, et il se peut que vous n'arriviez à en atteindre aucun de façon adéquate.

Les financeurs peuvent avoir un intérêt particulier pour l'efficacité et la rentabilité du programme, alors que le personnel du programme peut vouloir en savoir plus sur ce qui a fonctionné et échoué ainsi que sur la manière de mieux faire à l'avenir. Quant à la direction, elle peut s'intéresser aux implications politiques plus larges des conclusions de l'évaluation.



**Astuce :** Observez comment les parties prenantes indirectes peuvent être influencées ou encouragées pour garantir que les utilisateurs visés utilisent bien les conclusions de l'évaluation (par exemple, la direction).

Votre rapport d'évaluation doit expliciter qui sont vos principaux utilisateurs visés. Cela peut être très simple ou plus complexe, comme l'illustrent les deux exemples ci-dessous :

- « Les principaux utilisateurs de l'évaluation seront les parties prenantes au sein du ministère des Affaires étrangères et les partenaires de mise en œuvre de Danida, alors que le Parlement et le grand public sont susceptibles d'être les utilisateurs secondaires. » (Mowjee, Fleming et Toft, 2015: 16)
- « Cette évaluation présente de multiples parties prenantes. La première est le bureau de Concern Worldwide en République démocratique du Congo (RDC), afin d'alimenter leurs programmes en cours et à venir à Masisi. La seconde est le siège de Concern Worldwide, afin de contribuer à leur apprentissage à partir de la RDC et de ses potentielles applications dans d'autres contextes.  
Le troisième groupe de parties prenantes représente les bailleurs de Concern en RDC, c'est-à-dire OFDA [bureau de l'aide d'urgence en cas de catastrophe de l'USAID], qui finance cette évaluation, ECHO [bureau de l'aide humanitaire et de la protection civile de la Commission européenne] et Irish Aid, qui finance également les activités humanitaires de Concern au Nord Kivu. Enfin, l'évaluation peut être utile à d'autres acteurs de l'aide intéressés par des leçons pratiques et des données probantes en matière d'interventions de transfert monétaire. » (Bailey, 2013: 10)

Pour les évaluations de plus grande taille et plus complexes, il est possible que vous envisagiez d'analyser les parties prenantes en utilisant la traditionnelle matrice « pouvoir-intérêt », comme présenté dans le Schéma 3.1. Cette matrice a été construite durant la phase de démarrage de l'évaluation humanitaire inter-agences de la réponse à la crise en République centrafricaine entre 2013 et 2015. Elle a aidé l'équipe d'évaluation à comprendre les différentes parties prenantes de manière à les engager sans les surcharger et à promouvoir un engagement structuré des utilisateurs. En s'appuyant sur une analyse détaillée des parties prenantes de l'évaluation en fonction de leurs pouvoir et intérêt (voir Annexe 2 du rapport de démarrage), cette matrice proposait une stratégie d'engagement des utilisateurs aux responsables de l'évaluation, OCHA dans ce cas. A titre d'exemple, les évaluateurs cherchaient à « suivre » le gouvernement centrafricain et la population affectée, partant du principe que leur intérêt pour l'évaluation était faible (ils pourraient néanmoins bénéficier d'une opportunité pour apporter leurs points de vue, apprendre de la réponse, mesurer sa réussite, et alimenter en informations une stratégie future), et qu'ils disposaient d'un faible pouvoir pour ce qui est de contribuer au processus (les évaluateurs avaient néanmoins besoin de leur participation à la collecte des données et de leur acceptation formelle) (Lawday, 2015 ; communication personnelle avec Lawday, novembre 2015).

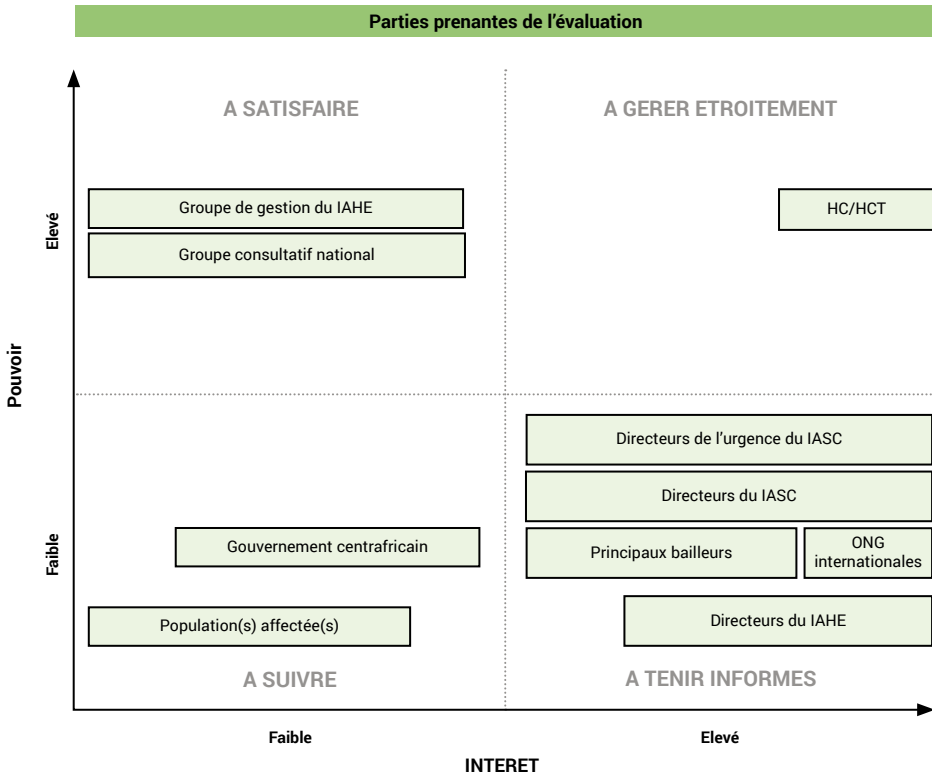
## Cartographie des parties prenantes

Différents groupes de parties prenantes peuvent être représentés en une série de cercles concentriques, en distinguant les principales parties prenantes qui sont censées être pleinement engagées dans l'évaluation, et celles qui ont un intérêt indirect et peuvent jouer un rôle différent. Celles qui ont un intérêt indirect sont également divisées entre celles qui devraient être consultées et influencées par l'évaluation, et celles qui devraient seulement être consultées.

Une manière de réaliser ce type de carte consiste à noter toutes les différentes parties prenantes sur des post-it, à dessiner des cercles sur un tableau à feuilles, et à placer chaque post-it sur le cercle approprié selon les intérêts des parties prenantes et leur degré d'engagement annoncé dans l'évaluation. Cela peut prendre la forme d'un exercice rapide et participatif, réalisable dans le cadre de petites évaluations.



Schéma 3.1 Analyse des parties prenantes de l'évaluation



Source : Lawday (2015) Pour plus d'informations sur cet outil, voir Start et Hovland (2004: 26).

### De quelles informations et données probantes ont besoin les utilisateurs ?

Une fois que vous avez identifié vos principales parties prenantes, demandez-leur : qu'avez-vous besoin de savoir pour mieux décider quoi faire et comment ? Généralement posée par le commanditaire ou le responsable de l'évaluation, cette question aide à clarifier et articuler les buts de l'évaluation.

Le périmètre d'une évaluation peut facilement devenir très large si vous demandez aux parties prenantes ce qu'elles *aimeraient* savoir, ce qui peut donner lieu à une liste ingérable de points à couvrir. Encouragez une focalisation sur l'utilisation en mettant l'accent sur la connaissance « pour l'action », cela en posant aux principaux utilisateurs visés des questions comme : qu'avez-vous *besoin* de savoir pour vraiment progresser ?

Les réponses doivent vous permettre de préciser l'objectif et le périmètre de l'évaluation. Elles doivent également vous éclairer sur la manière dont l'information et les données probantes collectées tout au long de l'évaluation peuvent être le mieux présentées pour faciliter leur utilisation. Par exemple, quel devrait être le format et/ou la longueur du rapport ? Quelles autres formes de diffusion fonctionneraient avec les utilisateurs visés ? Existe-t-il des réunions ou des forums de décision clés auxquels les conclusions de l'évaluation pourraient être présentées, cela de façon immédiatement pertinente ? Pour plus de détails sur ces options voir [Section 17 : Rédiger et communiquer des résultats d'évaluation en se focalisant sur l'utilisation.](#)

## Impliquer vos principaux utilisateurs visés

Les personnes auront davantage tendance à utiliser une évaluation si elles comprennent son processus et ses conclusions, et se les approprient, ce qui est également plus probable si elles ont été activement impliquées tout au long du processus (Patton, 1997). Cela signifie qu'il faut inclure des interactions fréquentes entre les utilisateurs visés d'une part et, d'autre part, les responsables et les évaluateurs de l'équipe. Cela demande un investissement en temps et en ressources.



### Question directrice

Pensez à votre processus d'évaluation. Quelles mesures ou activités spécifiques pourraient être rendues plus participatives ou inclure les utilisateurs ?

Voici quelques idées développées ultérieurement dans ce *Guide* :

- Impliquez les utilisateurs potentiels dans la compilation des [Termes de référence](#).
- Envisagez la possibilité d'inclure quelques utilisateurs dans l'[équipe d'évaluation](#). (Les organisations ont des politiques différentes à ce sujet, et certaines préfèrent impliquer des utilisateurs visés dans des groupes de référence externes plutôt que dans l'équipe d'évaluation.)
- Constituez un [groupe de référence](#) pour l'évaluation en incluant les principaux utilisateurs visés.
- Organisez un atelier où les principaux utilisateurs visés et l'équipe d'évaluation réalisent conjointement la conception de l'évaluation (voir [Exemple de bonnes pratiques, p. 72](#)).
- [Partagez la version provisoire du rapport](#) avec toutes les personnes qui ont été interrogées et donnez-leur la possibilité de le commenter.
- Impliquez les parties prenantes dans l'[analyse des données](#).
- Organisez un atelier pour [présenter les conclusions de l'évaluation](#).
- Demandez aux utilisateurs de participer à la [rédaction des recommandations](#).
- Demandez aux utilisateurs de concevoir une [stratégie de diffusion](#).

Il est également important d'envisager des activités qui couvrent l'ensemble du processus d'évaluation, par exemple :

- Assurez-vous que l'équipe d'évaluation communique de manière régulière et transparente avec les utilisateurs visés tout au long du processus d'évaluation. Cela aidera à maintenir les attentes et évitera les mauvaises surprises. Le fait de tenir informés les utilisateurs visés de l'avancée de l'évaluation et des changements éventuels de périmètre ou de méthodes facilitera, par exemple, l'utilisation des conclusions.
- Impliquez les principaux utilisateurs visés dans les décisions clés, par exemple à propos d'un recentrage de l'évaluation durant la phase de démarrage si les termes de référence (TdR) initiaux se sont avérés trop ambitieux pour les ressources disponibles. Les utilisateurs qui font partie du groupe de référence de l'évaluation peuvent contribuer à cela.

Les évaluations peuvent être une source d'inquiétude ou de résistance pour les personnes dont le travail est évalué. Ce phénomène peut être exacerbé dans les crises humanitaires à grande visibilité qui ont attiré l'attention des médias internationaux. Dans ces cas-là, il est d'autant plus important de construire un sentiment d'appropriation. Cela peut être équilibré avec le besoin de préserver l'objectivité ou l'indépendance de l'évaluation si les évaluateurs respectent des normes et des lignes directrices professionnelles. L'objectivité ne signifie pas s'éloigner des besoins des utilisateurs. Les évaluateurs doivent avoir les compétences interpersonnelles nécessaires pour établir des relations constructives avec les utilisateurs clés et construire de la confiance. Une communication fréquente tout au long de l'évaluation peut également aider à cela. Créer un engagement vis-à-vis de l'évaluation implique souvent de promouvoir une attitude d'ouverture au changement.



**Exemple de bonnes pratiques : Impliquer les parties prenantes clés tout au long de l'évaluation**

L'évaluation de la réponse d'Emergency Capacity Building Project au tremblement de terre de Yogyakarta (Wilson *et al.*, 2007) s'est fortement concentrée sur l'utilisation. Parmi les aspects clés, on retrouvait :

- Des discussions au début de l'évaluation avec les membres du comité de pilotage au niveau du pays par rapport à leurs attentes
- Un dialogue entre le chef d'équipe et le comité de pilotage avant que le chef d'équipe n'arrive dans le pays
- Des discussions entre l'équipe d'évaluation et le comité de pilotage à propos des méthodes d'évaluation - par exemple, l'interprétation la plus appropriée des critères du DAC.

**Exemple de bonnes pratiques : Aider les parties prenantes clés à clarifier leurs objectifs**

Parmi les exemples de parties prenantes travaillant ensemble pour définir les objectifs d'une évaluation, on peut citer celui de l'étude d'apprentissage conjointe qui portait sur la reconstruction de logements suite au tremblement de terre de Yogyakarta en 2006. Les organisations qui ont participé à cette étude étaient la Croix-Rouge suisse, Medair, Caritas Suisse, Solidar, Heks et Swiss Solidarity.

Avant l'évaluation, les utilisateurs visés, en particulier le personnel des organisations au sein des sièges respectifs, les évaluateurs, ebaix, et le bailleur, Swiss Solidarity, ont participé à un atelier animé par un consultant externe pour développer et identifier des sujets d'apprentissage clés. Des méthodes participatives, comme l'« open space », ont été utilisées afin de permettre aux participants de choisir des sujets d'apprentissage pertinents pour les programmes de reconstruction humanitaire en cours et à venir, ce qui a par la suite permis de définir les domaines à étudier. Les participants ont partagé des exemples de projet en se focalisant sur le travail de terrain (Otto, 2014).

### 3.4 Le rôle de la population affectée dans la phase de planification

Comme indiqué précédemment, les évaluations ne sont généralement pas la manière la plus appropriée pour rendre des comptes à la population affectée (voir [Section 14 : Impliquer la population affectée dans votre évaluation](#)). On trouve encore peu d'exemples de consultation de la population affectée en ce qui concerne l'angle de l'évaluation ou les questions à poser. [L'exemple de bonnes pratiques p 73](#) ci-dessous est un exemple intéressant mais rare de consultation réalisée lors d'une évaluation conjointe d'impact humanitaire (JHIE).





### Exemple de bonnes pratiques : Consultation de la population affectée à propos du potentiel de la JHIE en 2010

De manière à évaluer si la JHIE était soutenue, et afin d'explorer des approches possibles, des consultations ont été réalisées entre février et novembre 2010, notamment auprès de la population affectée dans 15 communautés au Bangladesh, en Haïti et au Soudan, mais aussi auprès d'acteurs humanitaires et de représentants du gouvernement dans chaque pays. « C'est peut-être la tentative la plus systématique de consultation des populations affectées durant la phase de conception d'un exercice évaluatif majeur » (Beck, 2011 : iv). Certains membres de la population affectée ont exprimé leur désir de participer activement au processus de l'évaluation et de recevoir des informations sur les résultats de la JHIE de manière à pouvoir jouer un rôle dans la validation des conclusions. Certaines personnes ont exprimé leur scepticisme sur la capacité d'une telle évaluation à apporter une réelle valeur ajoutée : « Comme l'a remarqué une communauté au Soudan du Sud, les gens étaient préoccupés par le fait qu'ils avaient participé à plusieurs évaluations, mais qu'il n'y avait pas eu de changements dans les erreurs qu'ils avaient identifiés » (*ibid*: iv).

Source : Beck (2011)

Si des consultations et du dialogue efficaces ont eu lieu entre les organisations et la population affectée tout au long de la mise en œuvre, alors les personnes en charge de planifier et de concevoir l'évaluation peuvent avoir une idée claire des types de sujets à explorer. A titre d'exemple, si la population affectée a soulevé des problématiques en matière de protection, alors il peut être approprié d'inclure une question évaluative sur l'efficacité avec laquelle l'organisation ou le programme a abordé la protection. Pour plus d'informations sur l'implication de la population affectée dans votre évaluation, voir [Section 14](#).

## 3.5 Les facteurs affectant l'utilisation des évaluations

« Le changement n'est pas un processus rationnel, mais un processus qui se produit au fil d'interactions variées et complexes au sein d'une organisation. De ce fait, exploiter toute la puissance de l'EAH en vue d'opérer un changement demande de faire attention à toute une gamme de sujets interconnectés » (Hallam et Bonino, 2013: 18-19). Cette sous-section présente deux cadres qui vous aideront à réfléchir aux facteurs qui influencent l'utilisation de l'évaluation. Le premier est adapté du cadre RAPID Context, Evidence, Links de l'Overseas Development Institute (ODI) servant à examiner les liens entre la recherche et les politiques (Crewe et Young, 2002) ; il peut être utilisé pour réfléchir au contexte plus général et à la manière de promouvoir l'utilisation de votre évaluation au sein de ce contexte.

Le second est le cadre des domaines de capacités de l'évaluation développé par ALNAP (Hallam et Bonino, 2013) et il se focalise sur l'utilisation de l'évaluation au niveau organisationnel.

## Contexte, données probantes, cadre de liens

**Qualité de l'évaluation** – La qualité de l'ensemble du processus d'évaluation, et pas seulement du produit final, constitue un facteur critique pour ce qui est de déterminer son utilité et la probabilité qu'elle soit utilisée. C'est le facteur que l'équipe d'évaluation peut le plus contrôler et qui a un impact important sur la crédibilité de l'évaluation. Comme le mentionnait Sandison dans sa conclusion : « L'image de l'utilisation qui ressort des [études sur le sujet] est complexe et souvent inattendue. L'une des rares certitudes réside dans le fait que la manière avec laquelle - et la raison pour laquelle - une évaluation est réalisée affectent de façon significative la probabilité qu'elle soit utilisée » (2006: 91).

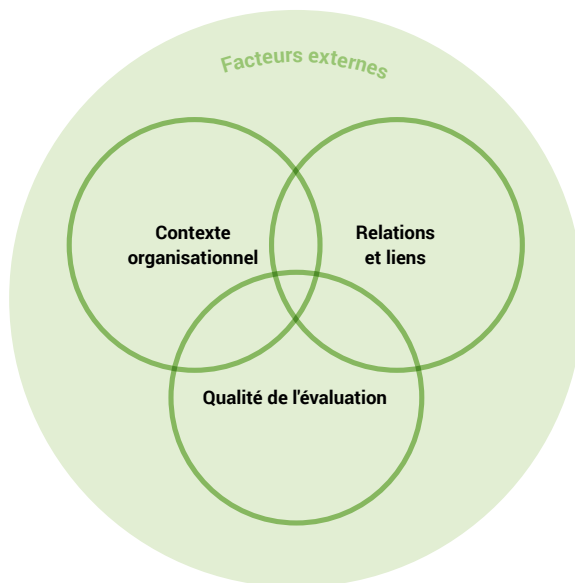
**Contexte organisationnel** – Cela comprend des sujets plus larges comme la culture de l'organisation en matière d'apprentissage et de gestion de la connaissance, ainsi que la structure organisationnelle en termes de proximité des unités d'évaluation vis-à-vis des décideurs. Les évaluations sont bien plus susceptibles d'être utilisées au sein d'une organisation qui recherche activement des informations sur sa performance de manière à améliorer la gestion et la mise en œuvre de ses programmes. Ce processus peut être fortement influencé, soit quand les principaux dirigeants soutiennent l'apprentissage, soit quand ils attachent peu d'importance à une prise de décision basée sur des données probantes et sont donc enclins à être sur la défensive vis-à-vis de l'évaluation.

**Relations et liens** – Cela comprend la relation entre les évaluateurs et les utilisateurs de l'évaluation. Le facteur personnel – par exemple, la présence d'un ou plusieurs « poids lourds » de l'évaluation parmi les parties prenantes clés avec lesquels il est difficile pour les évaluateurs d'établir et de maintenir une relation – peut être crucial pour déterminer si une évaluation est utilisée ou non. La position formelle et l'autorité de tels « poids lourds » peuvent avoir moins d'importance que leur enthousiasme et leur intérêt.

**Facteurs externes** – Cela peut comprendre le public ou les médias, par exemple dans une réponse humanitaire présentant une haute visibilité qui pose des questions sur la manière dont d'énormes sommes d'argent collectées auprès du grand public ont été utilisées. Ce fut par exemple le cas lors de la réponse au tsunami de 2006 dans l'océan Indien. Les « facteurs externes » peuvent également renvoyer au contexte plus large des politiques humanitaires et à l'ouverture, possible ou non, au changement. Certains observateurs pensent que le système humanitaire est « réactif au changement sous la pression lorsque les facteurs d'incitation sont à un niveau élevé » (Buchanan-Smith, 2005: 98), comme ce fut le cas lors de l'évaluation multi-organisations au Rwanda (voir [Exemple de bonnes pratiques p.75](#)). D'autre part, la synthèse du rapport de la Tsunami Evaluation Coalition (TEC) concluait que le manque de pression extérieure au niveau du changement est une raison clé de la si faible amélioration de la performance dans les dix années qui ont précédé la réponse au tsunami (Telford et Cosgrave, 2006).

Ces facteurs externes sont représentés sous la forme d'un modèle à trois cercles dans le Schéma 3.2, avec des circonstances externes représentées par la zone grisée autour des cercles. Au moment de planifier votre évaluation, prenez le temps d'observer comment chacun de ces cercles peut faciliter ou gêner l'utilisation des conclusions de l'évaluation, mais aussi ce que vous pouvez faire pour lever les blocages et promouvoir l'utilisation.

Schéma 3.2 Les facteurs affectant l'utilisation d'une EAH



Note : Pour plus d'éléments sur ce cadre, voir Start et Hovland (2004: 6).



#### **Exemple de bonnes pratiques : L'évaluation conjointe de l'aide d'urgence au Rwanda**

L'évaluation conjointe de l'aide d'urgence au Rwanda en 1996 (Borton *et al.* 1996) est l'une des évaluations les plus influentes de l'action humanitaire jamais entreprises (Buchanan-Smith, 2003).

Le pourquoi et le comment d'un tel phénomène ont été étudiés quelques années plus tard et sa réussite peut être expliquée à l'aide du modèle à trois cercles (Schéma 3.2) :

- Cette évaluation était de grande qualité, en raison de la minutie et de la rigueur du travail, des compétences de l'équipe, et de la manière dont elle a formulé ses recommandations. Cela lui a valu une grande crédibilité auprès des utilisateurs potentiels.





- En termes de culture et de structure organisationnelle, le système humanitaire international était ouvert au changement à ce moment-là, en partie à cause du malaise répandu concernant les performances très variables au sein des organisations humanitaires.
- Un grand nombre d'individus haut-placés ont défendu l'utilisation de l'évaluation et ont habilement su créer et exploiter des alliances ou des réseaux en leur nom.
- En termes de relations, les évaluateurs avaient des liens étroits avec les décideurs politiques du secteur.
- Les influences externes étaient toutes aussi importantes. Après le choc de la crise au Rwanda et son intense couverture médiatique, les décideurs politiques étaient plus ouverts au changement.

**Exemple de bonnes pratiques : L'évaluation de l'aide humanitaire de l'Allemagne en 2009 par le Bureau fédéral des Affaires étrangères et le ministère fédéral de la Coopération économique et du Développement**

En 2009, le Bureau fédéral des Affaires étrangères et le ministère fédéral de la Coopération économique et du Développement ont entrepris une évaluation indépendante de l'aide humanitaire de l'Allemagne de manière à en tirer des leçons susceptibles d'être utilisées à des fins de meilleure gestion. Ce processus a beaucoup attiré l'attention, parce que c'était la première fois que les deux ministères entreprenaient une évaluation ensemble mais aussi parce que c'était la première fois que l'aide humanitaire officielle de l'Allemagne étaient évaluées par des experts externes.

Achevée en 2011, cette évaluation a eu des effets significatifs sur l'architecture de l'aide humanitaire allemande, ce qui a été facilité par un certain nombre de facteurs. Tout d'abord, l'évaluation coïncidait avec une réforme majeure du système humanitaire allemand. Les nouveaux responsables désignés étaient impliqués dans les résultats de l'évaluation et ont utilisé ses conclusions dans le cadre du processus de réforme. Les conclusions ont alimenté la version initiale de la toute première stratégie humanitaire du bureau des Affaires étrangères qui a lancé une initiative sur la qualité sans précédent, laquelle a donné lieu à de nombreuses activités en lien avec des sujets traités dans l'évaluation.

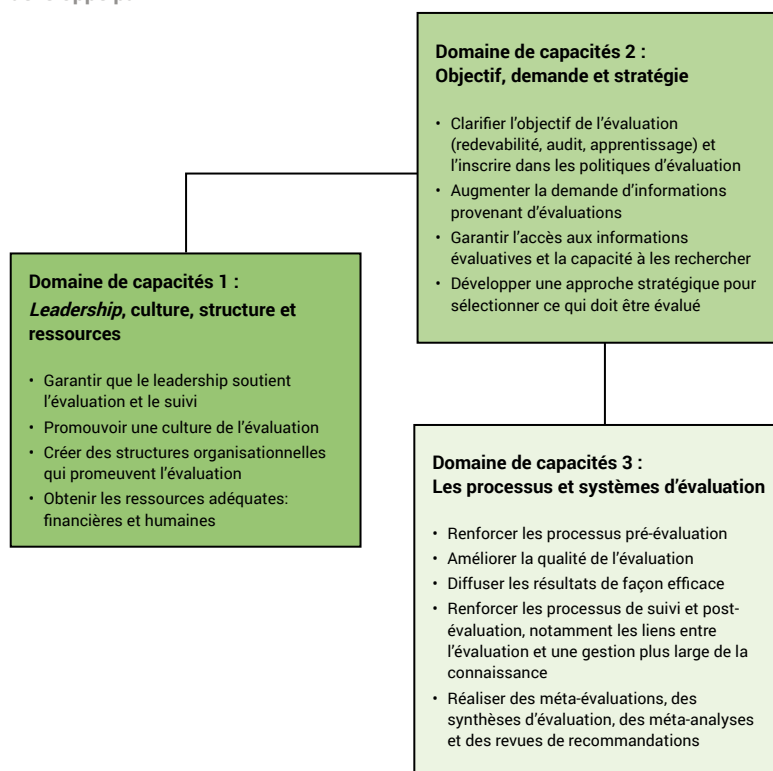
La recommandation de l'évaluation visant à prioriser les partenaires, les pays et les secteurs les plus solides demeure importante dans la manière dont l'aide humanitaire officielle est définie. Par exemple, le Bureau des Affaires étrangères recherche également des façons de mieux identifier les priorités en termes de secteurs et de pays.

Source : Weingärtner *et al.* (2011)

## Le cadre des domaines de capacités de l'évaluation développé par ALNAP

Hallam et Bonino (2013) présentent un cadre analytique simple pour faciliter l'analyse et la discussion des facteurs qui influencent la probabilité que des évaluations aboutissent à un impact au sein d'une organisation. Comme l'expliquent les auteurs : « La pensée évaluative doit être récompensée, et doit irriguer toute l'organisation. Toutefois, les personnes doivent désirer le changement, et être récompensées lorsqu'elles le mettent en œuvre si l'on veut que des améliorations se produisent dans leurs programmes. Elles doivent avoir les outils et le temps pour analyser ce qu'elles font, de même qu'elles doivent étudier les conclusions d'une manière qui encouragera le changement. L'organisation dans son ensemble doit accepter qu'en conséquence il y aura une remise en question des façons de travailler en place et un débat critique. Le *leadership* et la culture sont d'une importance cruciale, comme le sont également les systèmes formels et informels et les mécanismes de gratification ». (Hallam et Bonino, 2013: 18-19)

Schéma 3.3 Le cadre des domaines de capacités de l'évaluation développé par ALNAP



# 4 / Les types d'évaluation

Une fois que vous avez décidé qui sont les principaux utilisateurs visés, ce qu'ils veulent et ce qu'ils doivent savoir, ainsi que l'objectif global de l'évaluation en termes de redevabilité et/ou d'apprentissage, vous devez ensuite décider quel sera le type d'évaluation (FICR, 2010: 15-16).

## 4.1 Quel type d'évaluation allez-vous entreprendre ?

En identifiant le type d'évaluation que vous allez entreprendre, posez les questions suivantes :

1. **Quel est le périmètre de l'évaluation ?**  
Se concentre-t-elle sur le niveau projet, programme, secteur, thématique ou global ?
2. **Quel niveau de résultats vous intéresse ?**  
Voulez-vous évaluer le processus, les résultats ou l'impact? Jusqu'à quel niveau de détail voulez-vous observer la chaîne de résultats ? (Voir Section 5 : Elaborer votre évaluation)
3. **Quel est le *timing* de l'évaluation par rapport à la mise en œuvre ?**  
L'évaluation vise-t-elle à influencer la définition du programme à une phase avancée, comme dans le cas d'une évaluation en temps réel (ETR) ? Ou bien est-ce une évaluation à mi-parcours ? Ou alors intervient-elle après la fin du programme (*ex-post*) ? Le *timing* sera fortement influencé par l'objectif général de l'évaluation, selon qu'il est tourné vers l'apprentissage ou la redevabilité
4. **Combien d'acteurs sont impliqués dans l'évaluation?**  
S'agit-il d'une évaluation portant sur une seule organisation, d'une évaluation conjointe impliquant deux ou plusieurs acteurs, ou d'une évaluation portant sur l'ensemble du système humanitaire ?
5. **Qui est impliqué dans la mise en œuvre de l'évaluation ?**  
La plupart des EAH sont réalisées par des équipes qui n'ont pas été impliquées dans la mise en œuvre du programme ou par des équipes combinant évaluateurs externes et internes. Parfois, les personnes qui ont été impliquées dans la mise en œuvre du programme entreprennent une auto-évaluation. Les évaluations peuvent également être réalisées de façon participative et impliquer la population affectée du début à la fin, mais c'est extrêmement rare dans l'EAH.

## 6. L'évaluation contient-elle d'autres aspects remarquables ?

Par exemple, l'évaluation se concentre-t-elle sur les politiques ou s'agit-il d'une méta-évaluation ?

Ces catégories sont poreuses et les EAH relèvent en général de plusieurs catégories en même temps. Par exemple, une évaluation de programme (périmètre) peut être une évaluation d'impact (niveau des résultats), être conduite *ex-post* (niveau du *timing*), être une évaluation conjointe (impliquant plusieurs acteurs) réalisée de façon indépendante (niveau des parties prenantes qui la mènent à bien). Le type d'évaluation, ou la combinaison de différents types, aide également à déterminer comment concevoir l'évaluation et quel type de questions sera approprié.

Les réponses aux questions ci-dessus doivent clairement apparaître dans les termes de référence de l'évaluation (TdR).

## 4.2 Définir les différents types d'évaluation

### Les évaluations présentant des périmètres différents



#### Définition : Evaluation de projet

Evaluation d'une unique intervention humanitaire (avec des objectifs, des ressources et un calendrier de mise en œuvre spécifiques) qui se déroule souvent dans le cadre d'un programme plus vaste.

Un exemple illustrant ce type est l'évaluation des projets de construction d'abris de Tearfund et Tear Netherlands après le tremblement de terre de 2009 à Padang, sur l'île de Sumatra (Goyder, 2010). Parfois, un petit groupe de projets associés de la même organisation peut être évalué, comme dans le cas de projets financés par ECHO au Pakistan et mis en œuvre par le NRC (Murtaza *et al.*, 2013).



#### Définition : Evaluation de programme

Evaluation d'un ensemble d'interventions présentant un objectif humanitaire partagé.

Un exemple illustrant ce type est l'évaluation de l'aide fournie par Danida aux personnes déplacées internes en Angola (Cosgrave, 2004). Des programmes émergent parfois d'un ensemble d'interventions présentant des objectifs communs (comme dans cet exemple), sans avoir été conçus dès le départ comme un programme cohérent, comme dans le cas du programme d'éducation en Palestine du NRC (Shah, 2014).

**Définition : Evaluation de Cluster**

Evaluation de multiples projets au sein d'un programme plus vaste, OU évaluation en lien avec le système de coordination par Cluster des Nations unies.

Parmi les exemples d'évaluations de Cluster, on trouve l'évaluation réalisée par AusAid d'un cluster d'ONG après le tremblement de terre au Pakistan (Crawford *et al.*, 2006) et le travail d'ONG dans l'océan Pacifique (Crawford & Eagles, 2008).

Le terme « évaluation de Cluster » renvoie de plus en plus à l'évaluation du système de Coordination par Cluster des Nations unies, comme l'évaluation initiale réalisée par Stoddard *et al.*, 2007, l'évaluation du Cluster Logistique Global ou l'évaluation du Cluster Sécurité alimentaire Global (Steets *et al.*, 2013), ou encore du Cluster Nutrition en réponse aux inondations qui ont eu lieu au Pakistan en 2010 (Nutrition Cluster, 2011).

**Définition : Evaluation de partenaire**

Evaluation d'un ensemble d'interventions mises en œuvre par un seul partenaire.

L'évaluation Norad du travail de la NRC est un exemple de ce type (Ternström *et al.*, 2013), tout comme les évaluations que fait ECHO à propos de ses partenariats, par exemple avec l'Office de secours et de travaux des Nations unies pour les réfugiés de Palestine dans le Proche-Orient (Grünewald et de Geoffroy, 2009).

**Définition : Evaluation sectorielle**

Evaluation d'un groupe d'interventions dans un secteur, lesquelles contribuent toutes à atteindre un objectif humanitaire spécifique. L'évaluation peut couvrir une partie d'un pays, un pays entier, ou plusieurs pays (UNICEF, 2013).

L'évaluation du secteur eau, assainissement et hygiène en République démocratique du Congo est un des exemples d'évaluation sectorielle (Van der Wijk *et al.*, 2010). Pour certaines organisations, les évaluations sectorielles ont maintenant été remplacées par le second type d'évaluation de Cluster décrit ci-dessus, comme l'évaluation du Cluster Abris et Produits non alimentaires durant la réponse du FICR au cyclone tropical Giri au Myanmar en 2010 et 2011 (Hedlund, 2011).



**Définition : Evaluation thématique**

Evaluation portant sur une sélection d'interventions qui répondent toutes à une priorité humanitaire spécifique commune à plusieurs pays, régions, et éventuellement organisations et secteurs.

Parmi les exemples illustrant ce type d'évaluation, on trouve celle réalisée sur le rôle de l'aide alimentaire aux réfugiés en situation de déplacement prolongé (Cantelli *et al.*, 2012) ou encore l'évaluation du diagnostic des besoins suite au tsunami de décembre 2004 dans l'océan Indien (de Ville de Goyet et Morinière, 2006). Ce fut l'une des cinq études thématiques réalisées pour la Tsunami Evaluation Coalition. L'étude du HCR sur l'éducation des réfugiés constitue un exemple d'évaluation thématique portant sur un seul secteur (Dryden-Peterson, 2011).

**Définition : Evaluation d'un portfolio humanitaire**

Evaluation du portfolio humanitaire complet d'une organisation.

De telles évaluations sont conduites par des bailleurs afin de passer en revue l'ensemble de leur portfolio. Parmi les exemples, on trouve l'évaluation de l'action humanitaire finlandaise (Telford, 2005), celle de l'action humanitaire de la Suisse (de Ville de Goyet *et al.*, 2011) et celle de la stratégie humanitaire du Danemark (Mowjee *et al.*, 2015).

## Les évaluations à différents niveaux de résultats

**Définition : Evaluation d'impact**

Evaluation qui se concentre sur les effets plus larges d'un programme humanitaire, notamment l'impact intentionnel et non intentionnel, l'impact positif et négatif, et enfin l'impact macro (sectoriel) et micro (ménage, individu).

L'évaluation du Bureau de la Population, des Réfugiés et des Migrations du Département d'État américain portant sur l'aide aux réfugiés burundais (Telyukov *et al.*, 2009) constitue un bon exemple d'évaluation d'impact. Un autre excellent exemple est l'évaluation d'impact de la reconstruction du comté de Lofa au Liberia sur une initiative communautaire (Fearon *et al.*, 2008). De plus amples détails sur les évaluations d'impact sont disponibles à la [Section 18](#).

**Définition : Evaluation de processus**

Évaluation qui se concentre sur les processus par lesquels les moyens sont convertis en produits. Elle peut aussi examiner l'intervention dans son ensemble.

Les évaluations de processus ne cherchent pas à examiner les résultats ou l'impact mais plutôt les processus à l'œuvre « derrière » l'action humanitaire. A titre d'exemple, le HCR a évalué sa relation avec le Fonds central d'intervention d'urgence (CERF sous son acronyme anglais), ainsi que l'utilisation qu'il fait de ce dernier, afin d'étudier comment cela avait affecté le financement de ses programmes et contribué à des tendances de financement sur une période de cinq ans (Featherstone, 2014).

**Définition : Evaluation normative**

Évaluation qui compare ce qui est mis en œuvre avec ce qui était planifié ou avec des normes spécifiques.

Les évaluations normatives sont relativement rares dans l'EAH. Elles renvoient à un ensemble de normes et ne doivent pas être confondues avec des évaluations de travail normatif, lesquelles évaluent un travail dans le but d'établir des normes et des standards (UNEG, 2013). Un exemple d'évaluation normative est celle de l'évaluation de la réponse à la crise de 2002-2003 en Afrique australe par le Disasters Emergency Committee (DEC) (Cosgrave *et al.*, 2004) au regard du Code de conduite pour le Mouvement international de la Croix-Rouge et du Croissant-Rouge et pour les organisations non gouvernementales (ONG) lors des opérations de secours en cas de catastrophe (SCHR et FICR, 1994).

## Les évaluations à des moments différents

**Définition : Evaluation en temps réel (ETR)**

Évaluation d'une opération humanitaire en cours à mesure qu'elle se déroule.

Parmi les exemples d'évaluations en temps réel, on peut noter celle de la réponse du FICR au cyclone tropical Haiyan aux Philippines (Greenhaigh *et al.*, 2014), et celle de la réponse du HCR à la crise syrienne (Crisp *et al.*, 2013).

**Définition : Evaluation à mi-parcours**

Evaluation réalisée à la moitié d'une intervention.

Parmi les exemples d'évaluations à mi-parcours, on peut noter celle du projet de diagnostic des besoins en Syrie (Featherstone, 2013) et celle de la DG ECHO concernant la décision régionale « Sécheresse » en faveur de la Grande Corne de l'Afrique (Wilding *et al.*, 2009).

**Définition : Evaluation *ex-post***

Evaluation réalisée à la fin d'une intervention.

L'évaluation menée par CARE sur sa réponse d'urgence au tsunami dans deux districts du Sri Lanka (Bhattacharjee *et al.*, 2007) en est un exemple.

**Définition : Evaluation itérative**

Série d'évaluations conçue pour se dérouler tout au long d'une intervention.

On peut noter deux exemples d'évaluation itérative : la paire d'évaluations portant sur la réponse au tremblement de terre en Haïti menée par le IASC trois et vingt mois après le tremblement de terre (Grünewald *et al.*, 2010 ; Hidalgo et Théodate, 2012), ainsi qu'une évaluation du HCR qui s'est prolongée six ans après la fin d'une intervention (Skran, 2012).

**Définition : Evaluation *ex-ante***

Evaluation conduite avant qu'une intervention ne commence.

De telles évaluations, rares dans le secteur humanitaire, s'appuient sur les leçons apprises lors d'opérations similaires. C'est par exemple le cas de l'étude portant sur la mise en place d'un Corps volontaire européen d'aide humanitaire (Bruaene *et al.*, 2010).

## Les évaluations impliquant un nombre d'acteurs différent



### Définition : Evaluation au sein d'une même organisation

Evaluation menée par l'organisation qui a mis en place l'intervention.

Parmi les exemples de ce type courant d'EAH, on peut citer l'évaluation réalisée par CARE de sa réponse au tremblement de terre au Pakistan (Kirkby *et al.*, 2006) et l'évaluation des programmes de transferts monétaires au Soudan réalisée par Oxfam GB (Bush et Ati, 2007).



### Définition : Evaluation conjointe

Evaluation menée par deux organisations ou plus, évaluant le travail de deux organisations ou plus.

Les évaluations conjointes peuvent impliquer un bailleur et une organisation bénéficiaire, plusieurs organisations avec des mandats similaires, ou différents acteurs travaillant dans le même secteur. Parmi les exemples, on peut citer les évaluations conjointes réalisées par des ONG partenaires sur un projet d'urgence de renforcement des capacités après le tremblement de terre de Yogyakarta (Wilson *et al.*, 2007) et l'évaluation de Care et Save the Children portant sur leur réponse en Haïti (O'Hagen *et al.*, 2011). Parmi les évaluations conjointes de bailleurs, on peut citer l'évaluation conjointe de suivi sur les liens entre urgence, réhabilitation et développement (Brusset *et al.*, 2013). Enfin, parmi les exemples propres aux Nations unies, on peut noter l'évaluation conjointe PAM/HCR de la contribution de l'aide alimentaire aux réfugiés en situation de déplacement prolongé (Cantelli *et al.*, 2012).



### Définition : Evaluation de l'ensemble du système

Evaluation de la réponse du système humanitaire international à une crise humanitaire, ouverte à tous les acteurs du système.

On peut noter les deux exemples suivants : l'évaluation par la Tsunami Evaluation Coalition de la réponse internationale au tsunami de 2004 dans l'océan Indien (Telford *et al.*, 2006) et l'évaluation multi-organisations de la réponse à la crise au Rwanda (Borton *et al.*, 1996).

## Les évaluations réalisées par différentes parties prenantes

**Définition : Evaluations externes ou indépendantes**

Evaluation réalisées par des évaluateurs n'appartenant pas à l'équipe qui a mis en œuvre le programme.

Les évaluations externes ou indépendantes sont généralement une exigence des évaluations tournées vers la redevabilité parce qu'elles apportent un point de vue objectif.

**Définition : Auto-évaluation**

Evaluation réalisée par ceux qui conçoivent et mettent en œuvre une intervention, synonyme d'évaluation interne.

La plupart des auto-évaluations se focalisent sur l'apprentissage plutôt que sur la redevabilité, et donnent souvent lieu à des documents internes. L'un des exemples publiés est l'évaluation de la réponse de Medair au tsunami de 2004 au Sri Lanka. Cette évaluation fut réalisée par un membre de l'équipe initiale au Sri Lanka qui est retourné par la suite sur le terrain pour réaliser l'évaluation (Lee, 2005).

**Définition : Evaluation participative**

Evaluation durant laquelle les parties prenantes, y compris la population affectée, coopèrent pour concevoir, mener et interpréter une évaluation (voir également les éléments sur la conception participative à la Section 11).

Les évaluations participatives sont très rares dans le secteur humanitaire. King (2005) souligne une idée fausse courante selon laquelle le simple fait d'impliquer le personnel du programme, les bénéficiaires du programme ou les participants du programme de quelque manière que ce soit – le plus souvent en collectant des données auprès de ces acteurs – en fait une évaluation « participative ». Selon lui, le seul contact entre un évaluateur et des bénéficiaires du programme ou des participants n'est pas suffisant. Ce qui rend une évaluation « participative », c'est le rôle que les parties prenantes, et notamment ceux qui participent au programme, jouent et le degré de leur implication tout au long du processus d'évaluation (Alexander et Bonino, 2014).

## Les autres types d'évaluation



### Définition : Evaluation technologique

Evaluation d'une technique ou d'une technologie spécifique.

Ces évaluations ne sont pas courantes. Parmi les exemples, on peut citer l'examen de l'utilisation du Module d'assainissement à grande échelle de la Croix-Rouge britannique après le tremblement de terre de 2010 en Haïti (Fortune et Rasal, 2010) et l'évaluation du kit abri utilisé après les inondations de 2012 au Nigeria (Bravo *et al.*, 2014). De telles évaluations peuvent aussi porter sur des approches innovantes comme l'utilisation de l'approche communautaire pour gérer la malnutrition aiguë au Népal (Guerrero, 2010).



### Définition : Evaluation institutionnelle

Evaluation des dynamiques internes des organisations chargées de la mise en œuvre de l'aide, de leurs instruments politiques, de leurs mécanismes d'acheminement des services et de leurs pratiques de gestion ainsi que des liens entre ces éléments.

On peut citer en exemple les deux évaluations du programme de l'UNICEF pour le renforcement des capacités humanitaires qui ont été financées par le Département pour le Développement international du Royaume-Uni (DFID) (Bhattacharjee *et al.*, 2010 ; Brown *et al.*, 2005) et l'examen indépendant de la réponse opérationnelle de l'UNICEF au tremblement de terre de 2010 en Haïti (Bhattacharjee *et al.*, 2011).



### Définition : Evaluation des politiques

Evaluation qui examine les compréhensions, croyances et hypothèses qui rendent les projets individuels possibles et souhaitables. Elle peut évaluer à la fois l'efficacité d'une politique et la manière dont la politique a été mise en œuvre.

Les évaluations de politiques sont souvent d'assez grande ampleur et ont tendance à être davantage des examens que des évaluations, comme l'illustre l'examen par le HCR de sa politique relative aux réfugiés urbains à New Delhi (Morand et Crisp, 2013). Parfois, de tels examens des politiques couvrent l'ensemble du secteur humanitaire comme pour la *Revue de la réponse humanitaire des Nations unies* (Adinolfi *et al.*, 2005), et la *Revue de la réponse d'urgence humanitaire* de DFID (Ashdown, 2011).



### Définition : Méta-évaluation

Evaluation conçue afin d'assembler des résultats issus de plusieurs évaluations, ou évaluation analysant la qualité d'une ou plusieurs évaluation(s).

Parmi les exemples, on peut citer la méta-évaluation des évaluations conjointes réalisée par ALNAP (Beck and Buchanan-Smith, 2008). Scriven (2011) fournit une liste de vérification propre aux méta-évaluations du second type, dont la méta-évaluation de la qualité des évaluations de l'aide étrangère des Etats-Unis (Kumar et Eriksson, 2012) est un exemple, tout comme les rapports du Global Evaluation Reports Oversight System de l'UNICEF (Universalis, 2013). L'UNICEF a publié les standards à partir desquels ses évaluations sont mesurées (UNICEF Evaluation Office, 2013). Pour plus de détails sur les méta-évaluations, voir [Section 17.6](#).

## 4.3 Les évaluations en temps réel (ETR)



### Pour approfondir : Les évaluations en temps réel (ETR)

**Objectif principal** : Fournir, de manière participative et durant le travail de terrain, des retours à tous ceux qui sont impliqués dans la mise en œuvre et la gestion de la réponse humanitaire (Cosgrave *et al.*, 2009b).

**Utilisation** : Une très forte augmentation des ETR a été observée ces dernières années dans le secteur humanitaire car les organisations essaient de renforcer l'apport des évaluations à l'apprentissage, en particulier au niveau de l'apprentissage immédiat durant la mise en œuvre.

#### Les avantages des ETR :

1. **Opportunité/rapidité** – les ETR sont réalisées lorsque des décisions opérationnelles et politiques clés sont prises. Elles peuvent mettre en exergue des problèmes importants négligés au moment de répondre dans la précipitation aux besoins les plus urgents. Par exemple, l'ETR du Catholic Relief Service (CRS) au Pakistan en 2010 s'est déroulée seulement neuf semaines après le début de la réponse de l'organisation et comprenait un jour de réflexion avec l'équipe et les partenaires dans divers lieux. Au cours de cette journée, des plans d'action immédiate ont été élaborés (Hagens et Ishida, 2010).



2. **Faciliter l'interaction au sein du personnel et avec les évaluateurs** – les ETR impliquent un dialogue soutenu avec le personnel, à la fois sur le terrain et au siège. Elles peuvent servir de canal de communication entre ces deux types de personnel en contournant les obstacles bureaucratiques. Par exemple, l'ETR de 2010 portant sur la réponse du Comité permanent inter-organisations (IASC sous son acronyme anglais) aux inondations au Pakistan comprenait une seconde visite de terrain durant laquelle les résultats, conclusions et recommandations étaient discutés avec les parties prenantes de façon participative (Polastro *et al.*, 2011b).
3. **Perspective** – une équipe d'ETR peut aborder une urgence selon différents angles, en discutant avec le personnel à tous les échelons, dans différents pays, avec les populations affectées, ainsi qu'avec les organisations partenaires et les représentants du gouvernement. Cela peut apporter une vision de l'opération moins influencée par la résolution de problèmes quotidiens.

**Compétences des évaluateurs** : Les membres de l'équipe doivent avoir une expérience opérationnelle suffisante pour comprendre les produits, résultats et impacts les plus probables des actions en cours.

**Difficultés et solutions potentielles :**

- Finaliser le rapport de l'ETR avant d'avoir quitté le terrain, contrairement aux évaluations classiques.  
*Prenez en compte le temps de rédaction dans le calendrier du travail de terrain*
- Les principales parties prenantes des ETR sont l'équipe terrain et les personnes en charge des opérations au siège, la plupart d'entre eux travaillant sous forte pression avec peu de temps pour lire un rapport d'évaluation conventionnel.  
*Prévoyez du temps pour les débriefings sur le terrain dans le calendrier de votre évaluation, par exemple un travail sur le terrain à mi-parcours pour aborder des questions émergentes avec l'équipe terrain, et un débriefing final.*



## 4.4 Les évaluations conjointes



### Pour approfondir : L'avantages des évaluations conjointes<sup>1</sup>

**Objectif principal** : Evaluation inter-organisations visant à comprendre leur impact collectif, et/ou à promouvoir l'apprentissage au sein des organisations, et/ou à évaluer des aspects de la relation entre organisations. L'objectif déclaré des EAH conjointes est généralement double – redevabilité et apprentissage –, mais dans la pratique les organisations qui y participent, par exemple dans le cadre de la Tsunami Evaluation Coalition, les ont souvent trouvées plus utiles au niveau de l'apprentissage.

**Utilisation** : Les évaluations conjointes sont de plus en plus courantes dans le secteur de l'aide humanitaire, ce qui reflète une tendance plus globale à des approches conjointes de l'aide.<sup>1</sup> Les évaluations conjointes peuvent se rencontrer au niveau des circuits de financement multi-bailleurs, comme par exemple pour l'évaluation du soutien à la consolidation de la paix au Soudan du Sud (Bennett *et al.*, 2010) ; au niveau des consortiums d'ONG, comme par exemple l'évaluation multi-ONG de la réponse au tremblement de terre de Yogyakarta (Wilson *et al.*, 2007) et l'évaluation d'Action by Churches Together (ACT Alliance) de la réponse au tremblement de terre en Haïti (McGearty *et al.*, 2012) ; et au niveau du concept de « Transformative agenda » des Nations unies comme avec les évaluations en temps réel du IASC (Cosgrave *et al.*, 2010 ; Grünewald *et al.*, 2010 ; Polastro *et al.*, 2011a). Il est également possible de les retrouver dans les évaluations de niveau systémique, comme dans le cas du tsunami (Telford *et al.*, 2006).

#### Avantages :<sup>2</sup>

1. **Se concentrent sur une vision globale** – Les évaluations conjointes encouragent la focalisation sur la situation d'ensemble, ainsi que sur les questions stratégiques et politiques.
2. **Adaptées à l'évaluation d'impact** – Les évaluations conjointes évitent d'attribuer un impact à une organisation en particulier et peuvent être utilisées pour évaluer l'impact collectif de la réponse humanitaire.
3. **Meilleur bilan au niveau de la consultation de la population affectée** – Avec des ressources supérieures à celles de la plupart des évaluations conduites par une seule organisation, les évaluations conjointes ont un meilleur bilan au niveau de la consultation de la population affectée (Voir Beck et Buchanan-Smith, 2008).
4. **Utilisation** – Les évaluations conjointes bénéficient souvent d'une meilleure crédibilité et peuvent s'avérer utiles pour promouvoir le changement. Elles peuvent également remplir une fonction intéressante d'apprentissage au sein des organisations, aidant celles qui y participent à comprendre les approches des autres et à échanger leurs bonnes pratiques.



5. **Au niveau managérial et financier** – Les évaluations conjointes mettent en commun les ressources et les capacités de l'évaluation, et peuvent réduire les coûts de transaction si elles réduisent le nombre d'évaluations d'organisations individuelles consultant les mêmes parties prenantes, en particulier dans la zone affectée.

**Qualité des évaluateurs** : Les évaluations conjointes complexes demandent une combinaison de compétences techniques, politiques et interpersonnelles, particulièrement au niveau du chef d'équipe.

Le nombre d'évaluateurs disposant de telles compétences est limité. Anticiper la planification et le processus d'appel d'offres aidera à sécuriser les réponses de candidats qualifiés.

**Problèmes et solutions potentielles :**

- Etant donné le nombre de parties prenantes impliquées dans une évaluation conjointe, la négociation des TdR est susceptible de prendre beaucoup plus de temps que pour une évaluation d'organisation individuelle.  
*Il est important de prévoir davantage de temps pour la négociation des TdR. De fortes compétences de direction et d'animation sont essentielles pour cette phase de planification initiale de manière à gérer les négociations et à s'assurer que les TdR finaux sont clairs et concis, plutôt qu'une longue liste de questions non prioritaires.*
- La complexité des structures de gestion et les forts coûts de transaction associés aux évaluations conjointes peuvent constituer un obstacle à l'implication des plus petites organisations, par exemple les ONG de petite taille, et notamment les ONG nationales.  
*L'identification de modalités contractuelles adaptées aux ONG de petite taille, qui est moins exigeante que celle propre aux parties prenantes de plus grande taille – par exemple uniquement durant des moments clés de l'évaluation, et/ou avec des apports de fonds moindres que les plus grandes organisations – contribuera à garantir leur implication.*

# 5 / Elaborer votre évaluation

Cette section s'intéresse à la manière de construire votre évaluation. L'élaboration renvoie à la mise en place ou à l'identification d'une structure qui aidera à traduire les questions évaluatives générales en questions spécifiques. Cette traduction peut s'appuyer soit sur les hypothèses relatives à la manière dont l'intervention est supposée atteindre le changement souhaité (questions causales), soit sur les normes propres à de telles interventions (questions normatives).

Le choix d'un cadre causal ou normatif dépendra du contexte. Il est parfois plus approprié de concentrer son attention sur des questions normatives, par exemple lorsqu'il n'est pas possible de déterminer l'attribution et qu'il est difficile d'évaluer la contribution.

Cette élaboration consiste essentiellement à faire en sorte que les évaluateurs identifient les sujets les plus pertinents et se concentrent sur eux plutôt que sur d'autres moins importants pour permettre à l'intervention d'atteindre son objectif.

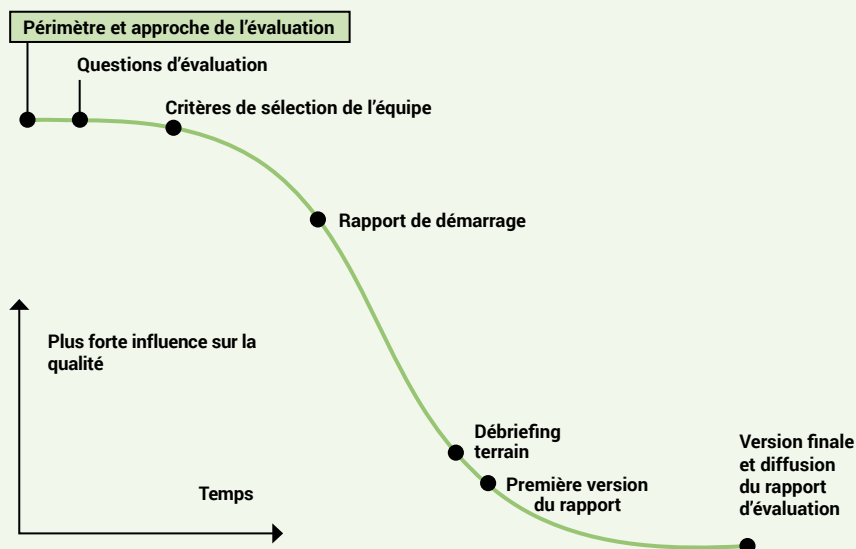
**Comment pouvez-vous garantir la qualité de votre évaluation ?** Une approche bien conçu et soigneusement structurée aide. Il y a plusieurs points clés pour le control de la qualité de l'évaluation. Trop souvent l'accent est mis sur le rapport d'évaluation pour le control de la qualité mais il vaut bien mieux se concentrer sur le control de la qualité en amont du processus.

Tout au long de ce guide nous soulignerons les points dans le processus d'évaluation durant lesquels les responsables de l'évaluation peuvent évaluer la qualité des produits et faire les ajustements pertinents.

Ces moments clés dans le control de la qualité sont représentés sur une courbe en S car le niveau d'influence qu'un responsable d'évaluation a sur la qualité diminue au fur et à mesure que le processus suit son cours.



## Point de contrôle de la qualité : Périmètre et approche de l'évaluation



Le premier point de contrôle clé de la qualité est le cadrage de votre évaluation, c'est-à-dire la définition du périmètre et de l'approche de votre évaluation.

## 5.1 La logique du programme

Durant la phase de conception des interventions, les actions planifiées doivent idéalement s'appuyer sur certaines théories concernant la manière dont elles atteindront le résultat final souhaité. Des théories explicites sont généralement présentées comme une sorte de modèle logique.



### Définition : Modèle logique

Tableau ou diagramme présentant la théorie du programme (manière dont les moyens sont supposés contribuer à l'objectif général d'une intervention).

Les modèles logiques vont des cadres logiques jusqu'aux diagrammes et aux théories du changement complexes (pour une présentation détaillée, voir Funnell et Rogers, 2011).

## Les cadres conceptuels

Les cadres conceptuels sont des modèles logiques, qui s'appuient en général sur une recherche approfondie, qui illustrent les facteurs clés d'un sujet en particulier. Cette base issue de la recherche signifie que les cadres conceptuels sont beaucoup plus robustes que les autres modèles logiques présentés ici. Alors que d'autres modèles logiques peuvent être fondés sur des liens de causalité supposés, les cadres conceptuels s'appuient sur des liens établis par des travaux de recherche.

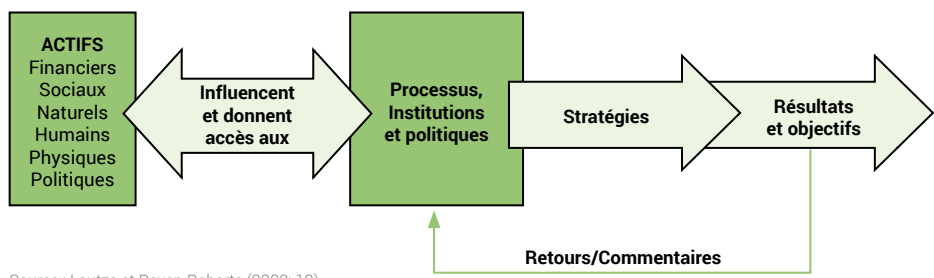


### Astuce :

Vérifiez si un cadre conceptuel couvre le sujet de votre évaluation. Les cadres conceptuels sont très utiles pour garantir qu'aucun domaine clé n'est négligé.

La cadre des moyens de subsistance durables s'intéresse aux moyens de subsistance ruraux. Lautze et Raven-Roberts (2003: 10) ont proposé la variante suivante du cadre pour des urgences humanitaires complexes.

Schéma 5.1 : L'outil analytique des moyens de subsistance



Source : Lautze et Raven-Roberts (2003: 10)

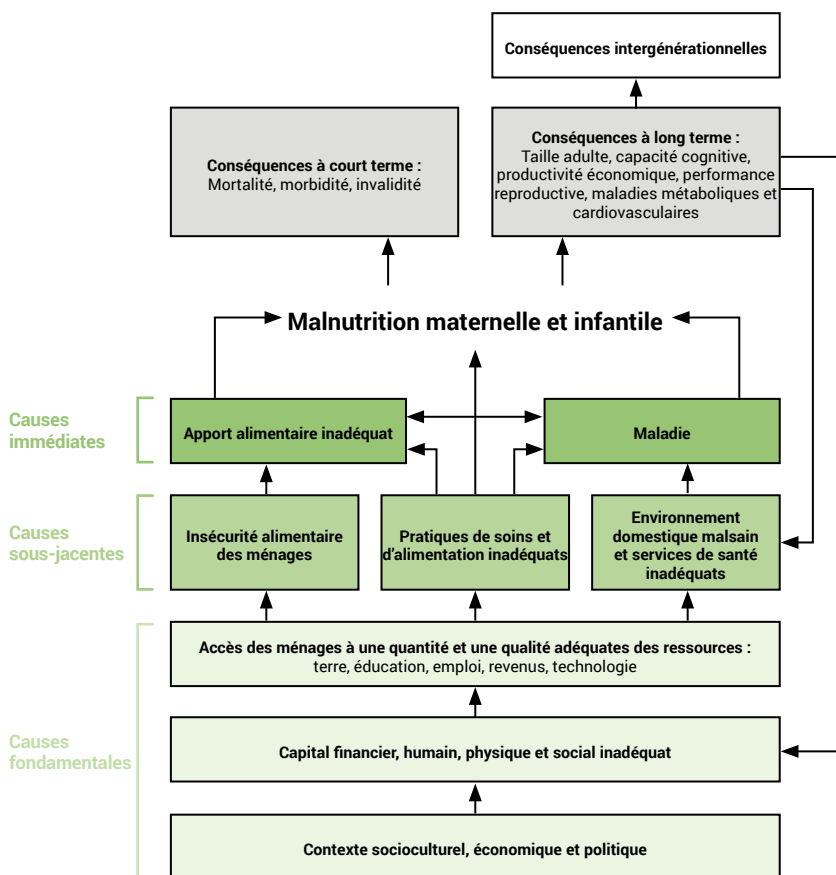
L'équipe de l'évaluation du programme de la FAO en Somalie a utilisé ce cadre (Buchanan-Smith *et al.*, 2013). Cela signifiait que tout en observant comment la FAO soutenait les moyens de subsistance via les actifs, notamment en soutenant les exploitations d'élevage par la distribution de médicaments et de fourrage, les évaluateurs s'intéressaient également à l'action de la FAO sur le contexte des politiques. Une question liée aux politiques était l'interdiction du bétail de la région par les pays du Moyen-Orient, ce qui avait une énorme influence sur les moyens de subsistance.

Un autre modèle conceptuel, qui peut être particulièrement utile lors d'évaluations de l'action humanitaire, est le modèle conceptuel de l'UNICEF en matière de malnutrition, surtout si l'accent de l'évaluation est mis sur les programmes d'aide alimentaire ou les interventions médicales.

L'évaluation du programme de sécurité alimentaire d'urgence et de moyens de subsistance urbains d'Oxfam GB (Macauslan et Phelps, 2012) a utilisé ce modèle conceptuel comme fondement du cadre causal de la malnutrition et de ses effets sur les moyens d'existence.

Même si aucun cadre conceptuel n'a été utilisé lors du développement d'une intervention, les évaluateurs peuvent néanmoins en utiliser un car il contient des liens de causalité avérés. Il peut également donner une idée des questions pertinentes. Par exemple, lors d'une évaluation d'un programme de nutrition, de mauvais résultats ne seraient pas surprenants si les questions de puériculture et de santé n'avaient pas été également traitées.

Schéma 5.2 : Le cadre conceptuel des déterminants de la malnutrition infantile de l'UNICEF



Source : UNICEF (2013).

## Les cadres logiques

Les cadres logiques sont l'une des formes les plus simples de modèle logique : les activités y sont directement liées aux résultats. Ce concept a été présenté pour la première fois dans le domaine du développement par USAID en 1970 (Rosenberg *et al.*, 1970) et a ensuite été plus largement adopté. Le Tableau 5.1 présente les éléments clés du cadre logique. Les organisations peuvent utiliser différents noms et définitions précises des éléments d'une chaîne de résultats.

Tableau 5.1 : Extrait de cadre logique

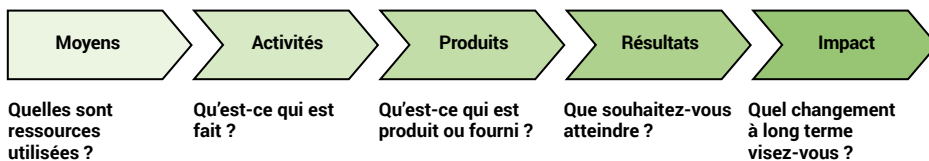
Hiérarchie des résultats	Indicateurs	Hypothèses
<b>Objectif général – l'objectif de niveau supérieur auquel le projet est censé contribuer</b>	Indicateurs mesurant l'atteinte de l'objectif général	Hypothèses concernant la durabilité de l'objectif général
<b>Objectif spécifique – l'effet qu'il est prévu d'atteindre comme résultat du projet</b>	Indicateurs mesurant l'atteinte de l'objectif spécifique	Hypothèses concernant l'atteinte de l'objectif général une fois l'objectif spécifique atteint
<b>Résultats – les résultats que les responsables du projet devraient être capables de garantir</b>	Indicateurs mesurant la proportion dans laquelle les résultats sont produits	Hypothèses concernant l'atteinte de l'objectif spécifique une fois les résultats en place
<b>Activités – les actions entreprises pour produire les résultats</b>	Moyens : les biens et services nécessaires pour entreprendre les activités	Hypothèses en lien avec la production des résultats

Source : Adapté de Norad (1999)

## La chaîne de résultats

L'approche du cadre logique a été critiquée pour sa rigidité excessive, en particulier dans des contextes humanitaires évoluant rapidement. De plus, la hiérarchie des résultats peut être source de confusion dans la mesure où l'objectif général d'un projet peut être le même que l'objectif spécifique d'un programme contenant ce projet.

Norad s'est finalement tourné vers une approche plus flexible en utilisant une chaîne des résultats comme indiqué ci-dessous.

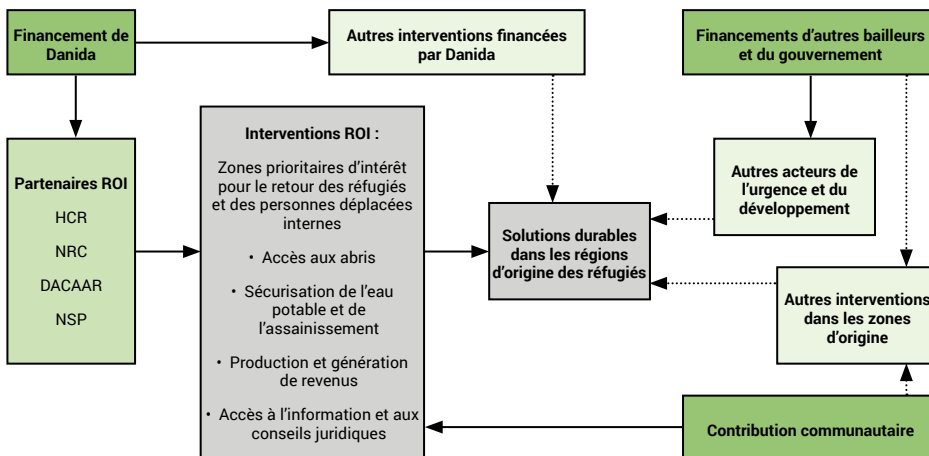


La chaîne de résultats peut avoir différents noms, y compris celui de parcours causal, mais l'idée centrale est de montrer comment les moyens sont censés mener au changement visé à long terme.

## Les modèles logiques graphiques

Les modèles peuvent également se présenter sous la forme d'un graphique, comme dans l'exemple ci-dessous d'une évaluation de l'aide danoise (ROI) en Afghanistan. Une telle présentation graphique peut aider à mettre en valeur les liens avec les actions d'autres organisations.

Schéma 5.3 : Evaluation de l'aide danoise en Afghanistan





La Fondation Kellogg a publié un guide sur le développement des modèles logiques (WK Kellogg Foundation, 2004) qui est construit autour de chaînes de résultats assez simples.



**Astuce :**

Les modèles sont des simplifications commodes d'une réalité complexe. Cette simplification est leur plus grande force car elle permet la manipulation de concepts et de situations complexes mais c'est aussi leur plus grande faiblesse car cela peut vous amener à négliger un facteur clé. Si vous utilisez un modèle lors d'une validation, examinez dans quelle mesure le modèle correspond à la situation réelle que vous examinez et modifiez-le en fonction.

Certains logiciels comme DoView peuvent faciliter le dessin d'un simple pipeline ou de modèles logiques de chaînes de résultats.

## 5.2 Les théories du changement



**Définition : Théorie du changement**

Une théorie du changement est une description du mécanisme central par lequel le changement se produit pour les individus, les groupes et les communautés.

Une théorie du changement (Tdc) exprime les liens de cause à effet ou les chaînes de liens entre l'intervention et le résultat final souhaité. C'est une approche de plus en plus populaire, et certains bailleurs ont adopté des Tdc à la place de modèles logiques (même si certains demandent les deux). Vogel remarque qu'elles sont perçues comme un « outil de réflexion plus réaliste et flexible que les approches cadre logique actuelles » (2012: 9). Le guide Tdc de Keystone compare Tdc et cadre logique (2009, 30), alors que Rogers (2012: 7) liste les cadres logiques comme l'une des quatre manières de représenter une Tdc.

Les théories du changement sont plus utiles dans des environnements complexes parce qu'elles permettent de décomposer la chaîne de résultats en une série de liens de causalité qui peut être testée par l'évaluation.

Toutefois, la Tdc fait encore moins consensus que d'autres formes de modèle logique. Stein et Valters remarquent (2012: 5) qu'il « existe un problème de fond dans le fait que différentes organisations utilisent le terme Tdc pour désigner des choses très différentes ».

Une TdC peut être représentée sous la forme d'un énoncé (pour des exemples, voir OECD/DAC, 2012: 85), d'un tableau ou d'un diagramme. L'exemple suivant, tiré d'une évaluation inter-organisations de l'aide alimentaire aux réfugiés de longue durée au Bangladesh, montre une présentation tabulaire d'une TdC comme « modèle logique simplifié » (Nielsen, 2012: 4).

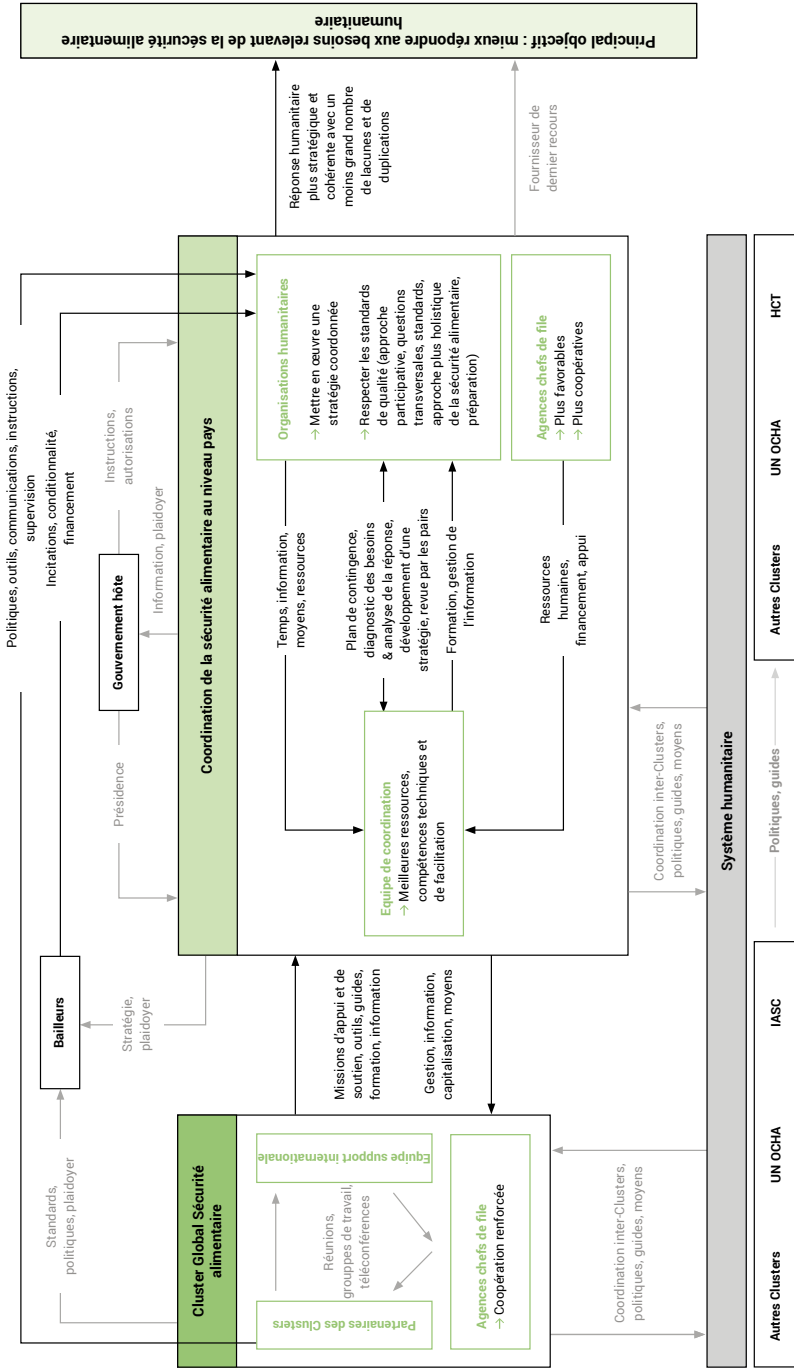
**Tableau 5.2** : Exemple d'une évaluation inter-organisations de l'aide alimentaire aux réfugiés de longue durée

Temps	Aide alimentaire	Hypothèses	Résultats
<b>Court terme</b>	Distribution alimentaire générale – rations complètes	Aide en réponse d'urgence	Vies sauvées ; consommation alimentaire améliorée ; sécurité et protection assurées. Niveau minimum d'autonomie.
<b>Moyen terme</b>	Diminution de l'aide alimentaire (rations partielles)	Transition pour sortir de la réponse d'urgence ; des interventions en services sociaux complémentaires sont disponibles, par exemple en eau, assainissement, éducation, logement, etc.	Panier alimentaire amélioré, état nutritionnel amélioré (malnutrition aigüe et chronique). Capacité accrue de la population affectée à mettre en place des moyens de subsistance.
<b>Long terme</b>	Diminution de l'aide alimentaire (rations partielles)	Des interventions liées aux moyens de subsistance sont disponibles ; développement des actifs.	Autonomie des réfugiés ; intégration locale ; réinstallation ou rapatriement.

Les auteurs du rapport de démarrage de l'évaluation du Cluster Sécurité alimentaire ont développé une TdC graphique (Steets *et al.*, 2013: 6) comme montré dans le [Schéma 5.4](#) de la page précédente.

Parmi les ressources disponibles sur le développement d'une TdC, on peut citer Funnell et Rogers (2011) ou encore Care International (2012). Par ailleurs, l'étude des TdC qu'a réalisée Comic Relief offre un aperçu utile. Les annexes du guide de l'OCDE-CAD portant sur l'évaluation des interventions de construction de la paix comprennent une section sur la compréhension et l'évaluation des TdC (OCDE/CAD, 2012: 80-86).

**Schéma 5.4 : Cluster Sécurité alimentaire - Théorie du changement**



## 5.3 Développer un modèle logique *a posteriori*

Idéalement, l'intervention du programme doit avoir été conçue autour d'une TdC de manière à produire un énoncé clair sur la manière dont elle doit fonctionner. Cependant, ce n'est pas encore une pratique commune, et le résultat final est que les évaluateurs peuvent se voir demander de développer dans les TdR une TdC explicite qui corresponde avec celle implicite de l'intervention.

En principe, une TdC implicite soutenait les décisions dans le cas du modèle logique ROI de Danida décrit plus haut, mais ne pas expliciter la TdC peut créer des problèmes :

- Une évaluation couvrant plus d'un seul projet ou programme est peu susceptible d'avoir une seule logique générale. Ce fut le cas de l'évaluation de la stratégie danoise en matière d'évaluation humanitaire (Mowjee *et al.*, 2015: 18).
- L'un des aspects les plus utiles d'une TdC explicite est qu'elle fournit un cadre permettant aux évaluateurs de tester la mise en œuvre. Si les évaluateurs développent la TdC, il existe un risque que les mêmes données soient utilisées à la fois pour générer la théorie et pour la tester.
- Si les évaluateurs trouvent qu'un lien de causalité particulier (comme supposé dans la TdC qu'ils ont développée) n'est pas valide, cela pourrait être dû à leur modèle plutôt qu'à l'intervention.

Une TdC *ex-post* peut n'être qu'un pâle ersatz de celle développée avant le début du programme et utilisée pour le guider. Développer une TdC après les actions relève plutôt de la tentative de réaliser une planification stratégique *ex-post*. Lorsque vous devez développer une théorie du changement *ex-post*, une approche consiste à organiser un atelier avec les parties prenantes pour définir la TdC implicite avec laquelle ils fonctionnaient ou à valider une TdC explicite que les évaluateurs ont définie grâce à la recherche documentaire.

## 5.4 Les cadres d'évaluation et les critères d'évaluation

Le second type d'élaboration passe par les normes de l'action humanitaire. Les contextes humanitaires sont souvent complexes, et le danger existe qu'un aspect essentiel soit négligé, que l'évaluation manque de rigueur conceptuelle ou qu'elle ne respecte pas les standards reconnus. Les cadres normatifs peuvent aider sur ces points des façons suivantes :

- Ils apportent une structure qui permet de diviser l'évaluation en de plus petits éléments plus faciles à gérer.
- Ils réduisent le risque que des éléments clés soient négligés, en dirigeant systématiquement l'attention de l'équipe vers tous les éléments du sujet de l'évaluation.
- Ils confèrent à l'évaluation une structure qui sera reconnue par les parties prenantes. Celles-ci peuvent avoir été très peu habituées aux critères standards de l'évaluation, mais elles sont souvent à l'aise avec la structure des cadres conceptuels ou des standards utilisés couramment dans leur secteur.
- Ils peuvent constituer un niveau de référence en matière de bonnes pratiques généralement reconnues, niveau par rapport auquel le projet peut être testé.

Les critères OCDE/CAD décrits dans la [Section 6 : Choisir les questions d'évaluation](#) fournissent un cadre possible pour une évaluation, mais tous les cadres se révèlent meilleurs pour vérifier que pour développer les questions.

D'autres cadres peuvent ajouter des détails et des spécificités à l'évaluation et la rendre ainsi plus accessible à ses utilisateurs. De tels cadres peuvent être utilisés en complément des critères OCDE/CAD, ou même se substituer à eux si nécessaire. On demande souvent aux évaluateurs d'utiliser au moins l'un des types de cadre suivants :

- De larges cadres normatifs qui reflètent les normes définissant l'engagement humanitaire. Ils comprennent le droit international humanitaire, les principes humanitaires (UNEG HEIG, 2016) et diverses conventions.
- Standards et guides peuvent être utilisés à la fois comme des standards pour l'évaluation et comme une manière de décomposer les actions humanitaires en de petits éléments plus faciles à examiner. On peut notamment citer les normes concernant l'ensemble du système et les normes sectorielles, mais aussi les guides et les manuels des organisations.

Les cadres et leurs utilisations sont résumés dans le [Tableau 5.3](#). Cette liste n'est pas exhaustive et de nouveaux cadres sont sans cesse en train d'être développés. Le cadre choisi doit être adapté au contexte.

Tableau 5.3 : Exemples de cadres et de leur utilisation

Cadres et exemples	Utilisation possible
<p><b>Les grands cadres normatifs</b></p> <ul style="list-style-type: none"> <li>• Droit humanitaire international comme le Protocole sur la protection des victimes de conflits armés non internationaux (CICR, 1977)</li> <li>• Principes humanitaires (Wortel, 2009)</li> <li>• Convention relative aux droits de l'enfant (Assemblée générale des Nations unies, 1989)</li> <li>• Convention relative au statut des réfugiés (Assemblée générale des Nations unies, 1951)</li> </ul>	<p>Les cadres normatifs peuvent être utilisés durant la phase de démarrage pour voir si des aspects de l'intervention pourraient susciter des inquiétudes et devraient être examinés de plus près. Certains cadres fournissent également des listes de vérification ou des standards avec lesquels examiner les politiques et l'efficacité.</p>
<p><b>Standards et lignes directrices concernant l'ensemble du système</b></p> <ul style="list-style-type: none"> <li>• Principes et bonnes pratiques pour l'aide humanitaire (Good Humanitarian Donorship, 2003)</li> <li>• Code de conduite des ONG et du Mouvement international Croix-Rouge/Croissant-Rouge (Borton, 1994)</li> <li>• La Norme humanitaire fondamentale sur la qualité et la redevabilité (CHS Alliance, Groupe URD et Sphere Project, 2014)</li> <li>• Principes directeurs relatifs au déplacement de personnes à l'intérieur de leur propre pays (OCHA, 2004)</li> <li>• Principes directeurs relatifs aux Etats fragiles (OCDE/CAD, 2007)</li> <li>• Compas qualité (Groupe URD, 2009)</li> </ul>	<p>Ces normes peuvent servir de liste de vérification ou de référence pour évaluer la performance, de base pour décomposer l'évaluation en tâches gérables, et de structure pour le rapport. Elles sont plus efficaces lorsqu'une organisation s'est formellement engagée à y adhérer ; si ce n'est pas le cas, leur utilisation peut être difficile à justifier. La phase de démarrage est le meilleur moment pour proposer l'utilisation d'une norme particulière.</p>
<p><b>Standards sectoriels</b></p> <ul style="list-style-type: none"> <li>• Eléments sectoriels des standards Sphère (Sphère Project, 2011)</li> <li>• Standards minimum de l'INEE pour l'éducation en situations d'urgence (INEE, 2006)</li> <li>• Directives et normes pour la gestion du bétail dans les situations d'urgence (LEGS Project, 2009)</li> <li>• Normes minimales pour le redressement économique (the SEEP Network, 2010)</li> <li>• Normes minimales pour la protection de l'enfance en situation humanitaire (Child Protection Working Group, 2012)</li> </ul>	<p>Les standards sectoriels représentent une bonne base pour organiser des évaluations sectorielles. Dans certains cas, ils s'appuient sur des standards généraux ; par exemple, le standard Sphère sur la consultation des populations affectées est censé s'appliquer à tous les secteurs.</p>



Cadres et exemples	Utilisation possible
<p><b>Standards et guides des organisations</b></p> <ul style="list-style-type: none"> <li>• Le manuel du Haut-Commissariat des Nations unies pour les réfugiés (UNHCR, 2007)</li> <li>• Le manuel du Programme alimentaire mondial relatif aux opérations d'urgence sur le terrain (PAM, 2002)</li> <li>• Le manuel de l'UNICEF relatif aux situations d'urgence (UNICEF, 2005)</li> </ul>	<p>Ces documents peuvent fournir une bonne base pour vérifier la conformité (redevabilité) et également pour décomposer et organiser le travail d'évaluation.</p>
<p><b>Standards et guides locaux</b></p> <ul style="list-style-type: none"> <li>• Lois, normes et politiques nationales</li> <li>• Lignes directrices reconnues localement (par exemple au niveau des Clusters)</li> </ul>	<p>Ces documents s'appuient souvent sur des modèles internationaux plus vastes, mais sont adaptés au contexte local. Ils sont ainsi plus appropriés en général au contexte que des normes et des guides internationaux plus vastes.</p>



**Astuce :**

Référez-vous aux documents de TdR et de projet. Si les documents de planification des TdR ou du projet renvoient à des cadres conceptuels ou des normes spécifiques, il peut être pertinent d'en utiliser un(e) comme cadre pour l'évaluation.

Il se peut qu'un problème survienne lorsqu'un rapport d'évaluation essaie d'utiliser plusieurs cadres à la fois. Cela peut donner lieu à un long rapport car chaque section du cadre A a des sous-sections pour le cadre B (et parfois même le cadre C).



**Astuce :**

Utilisez un seul cadre pour structurer le rapport. Si une évaluation utilise un cadre et que cela s'avère un élément central de l'évaluation, il peut fournir une structure utile pour le rapport. Si l'évaluation utilise plusieurs normes, nous recommandons d'en choisir un pour le cadre principal et d'aborder les autres dans de courtes annexes du rapport principal.

# 6 / Choisir les questions d'évaluation

Cette section se concentre sur le choix des questions d'évaluation. Celles-ci doivent s'appuyer sur ce que les principaux utilisateurs visés doivent savoir pour améliorer leur travail, comme expliqué dans la [Section 3 : Penser tôt et souvent à l'utilisation de l'évaluation](#). Les questions peuvent être divisées de la façon suivante :

- Les grandes questions, comme « Dans quelle mesure votre réponse a-t-elle été efficace » ?
- Les questions d'évaluation réelles, développées à partir des grandes questions
- Les questions posées aux personnes interrogées, aux *focus groups*, et aux personnes sondées dans les entretiens et les outils d'enquête.

Cette dernière catégorie est traitée dans la section sur la [Section 13 : Les méthodes de terrain](#).

## 6.1 Le choix des questions d'évaluation a un effet critique sur la qualité

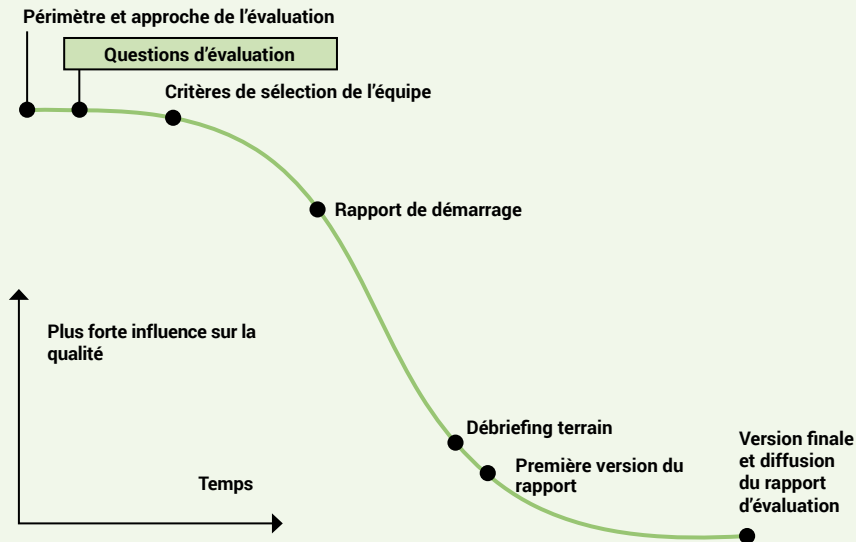
Le choix des questions d'évaluation a un effet critique sur la qualité de l'évaluation.

Les facteurs suivants peuvent donner lieu à une mauvaise évaluation :

- Trop de questions empêcheront l'évaluation d'aller en profondeur.
- Des questions qui correspondent mal avec le périmètre et l'approche.
- Des questions qui obtiennent de meilleures réponses par d'autres moyens (comme un audit ou une étude thématique).
- Des questions qui débouchent sur des réponses inutilisables.



## Point de contrôle de la qualité : Les questions de l'évaluation



Les questions que vous posez déterminent quelles sont les conceptions et les méthodes de recherche possibles. Les questions déterminent aussi quel budget sera nécessaire. De façon cruciale, le choix des questions a une forte influence sur l'utilisation de l'évaluation.

### Toutes les évaluations ne posent pas de questions

L'évaluation participative du programme Farmer Field School mis en œuvre à Lira en Ouganda (2008) ne comprenait pas de questions, mais seulement des objectifs d'évaluation. L'évaluateur a ensuite élaboré un tableau des informations obligatoires en lien avec les données nécessaires pour atteindre les objectifs (Foley, 2009). Dans le cas de l'évaluation de l'engagement danois en Somalie, aucune question n'a été communiquée aux évaluateurs à qui il a été demandé d'« approfondir et de spécifier la question d'évaluation relative aux sujets présentés plus haut » dans une matrice d'évaluation (Gartner *et al.*, 2012: 72). Même si cette approche est encore relativement rare, l'UE l'a désormais adoptée comme une pratique standard pour ses évaluations.

## Les questions d'évaluation orientent l'évaluation dans son ensemble

Les questions d'évaluation jouent un rôle critique dans la définition de l'évaluation. Cela vaut la peine de trouver les bonnes questions de manière à garantir la qualité de l'évaluation, car :

- Elles orientent le type d'évaluation choisie, par exemple l'évaluation d'impact, l'évaluation de processus ou l'ETR.
- Elles déterminent la ou les conception(s) la/les plus appropriée(s). Par exemple, si la question porte sur l'effet de l'aide, il se peut que vous ayez à réaliser une étude comparative comme ce fut le cas pour les études d'évaluation d'impact conjointes de l'aide alimentaire et l'étude de cas portant sur les crises prolongées de réfugiés au Rwanda (Sutter *et al.*, 2012).
- Elles déterminent quelles méthodes doivent être utilisées. Par exemple, si les TdR exigent que l'évaluation « détermine la satisfaction globale des bénéficiaires », il peut être nécessaire de réaliser une enquête, comme dans le cas de l'évaluation FICR des abris temporaires en Indonésie (Tango, 2009: 38).
- Elles orientent les options dans la matrice d'évaluation (discuté dans la Section 8 : La phase de démarrage. Voir Buchanan *et al.* (2009: 37-39) pour un exemple illustrant comment les questions donnent forme à la matrice d'évaluation.
- Les questions doivent orienter le budget de l'évaluation. Par exemple, les questions auxquelles on peut répondre seulement au moyen d'une enquête augmenteront le coût.
- Elles influencent à la fois le type et l'étendue des choix d'échantillonnage. La revue de l'initiative de l'UNICEF portant sur l'éducation en situation d'urgence et de transition post-crise a utilisé une stratégie d'échantillonnage par choix raisonné pour atteindre les objectifs de l'étude (Barakat *et al.*, 2010: 194).
- Les questions d'évaluation déterminent les recommandations. Comme le remarque le guide de l'OCDE/CAD sur l'évaluation de l'aide au développement : « Les résultats et les conclusions de l'évaluation sont les réponses aux questions posées et sélectionnées pour l'évaluation. Les leçons apprises et les recommandations créent le lien entre les résultats de l'évaluation et le développement des politiques et programmes à venir » (1991: 10).

## 6.2 Les types de questions d'évaluation

Le Tableau 6.1 présente cinq types de questions d'évaluation qui s'inspirent à la fois de BetterEvaluation (2013) et de Morra Imas et Rist (2009: 223-229).

Tableau 6.1 : Les types de questions d'évaluation

Type de question	Exemple	Commentaire
Descriptive	Comment les populations affectées ont-elles utilisé les kits « abri » fournis ?	Ce type de question est assez courant dans les EAH. Mais que faites-vous des résultats ?
Normative	Dans quelle mesure l'abri fourni respectait-il les standards Sphère ?	D'autres questions normatives peuvent approfondir la question de l'atteinte des cibles du projet ou des normes des organisations, ou la comparaison avec d'autres organisations. Elles se focalisent sur la <u>redevabilité plutôt que sur l'apprentissage</u> .
Causale	Dans quelle mesure la fourniture de l'aide au niveau du village a-t-elle découragé la migration vers la capitale régionale ?	Ce sont souvent les réponses les plus difficiles à formuler dans l'EAH car il est difficile d'évaluer l'attribution ou même la contribution étant donné la nature chaotique, complexe et compliquée de l'action humanitaire. Les questions d' <u>impact</u> sont causales par définition. Voir les éléments sur la causalité dans l'analyse à la <u>Section 16 : L'Analyse</u> .
Evaluative	Notre politique visant à ne fournir des kits « abri » qu'aux personnes ayant des preuves de propriété était-elle pertinente ?	Ce type de question demande aux évaluateurs d'émettre un jugement sur la valeur d'un programme. Voir les éléments sur la réflexion dans la <u>Section 16 : L'Analyse</u> . « Par définition, l'évaluation répond à des questions évaluatives, c'est-à-dire des questions portant sur la qualité et la valeur » (Davidson, 2014: 1). Cependant, la plupart des EAH vont au-delà et sollicitent des recommandations : les jugements de valeur ne donnent pas lieu, par eux-mêmes, à des recommandations.
Orientée vers l'action	Comment pourrions-nous mieux soutenir les personnes vulnérables pour reconstruire leur abri ?	Les questions orientées vers l'action conviennent particulièrement aux EAH car elles relient directement la question aux actions que l'organisation pourrait entreprendre à l'avenir.

Les questions orientées vers l'action sont complexes, et l'équipe d'évaluation doit généralement répondre à plusieurs sous-questions pour y répondre. Dans l'exemple donné sur les options permettant de mieux soutenir les personnes vulnérables dans le but de reconstruire leur abri, les sous-questions implicites sont notamment :

- Quelles sont les caractéristiques du projet ? (Descriptive)
- Qui rentre ici dans la catégorie « vulnérable » ? (Normative)
- Quel est le meilleur soutien ? (Evaluative)
- Quels facteurs permettent un meilleur soutien ? (Causale)

C'est seulement après avoir répondu à ces questions que l'équipe d'évaluation sera en mesure de lister les options et de faire des recommandations.



**Astuce :**

Là où cela est possible, formuler vos questions d'évaluation globales comme des questions orientées vers l'action.

Les questions orientées vers l'action conviennent bien à l'EAH parce que, comme les sous-questions pertinentes, elles se concentrent fortement sur la manière dont les conclusions seront réellement utilisées. Les questions orientées vers l'action permettent également d'excellents choix pour ce qui est de formuler les questions globales.

Le processus par lequel les évaluateurs répondent à des questions descriptives, normatives, causales et évaluatives puis formulent des recommandations est généralement un peu désordonné. Poser des questions orientées vers l'action signifie que les recommandations ne sont pas laissées au hasard ou aux préférences personnelles des évaluateurs, mais indiquent clairement où vous voulez qu'elles vous guident pour la future programmation.



**Astuce :**

Envisagez de demander aux évaluateurs des options plutôt que de simples recommandations.

La prise de décision relève de la direction. Le fait de demander des options implique que les évaluateurs étudient leurs avantages et inconvénients respectifs plutôt qu'ils ne formulent une simple recommandation. Cela n'empêche pas les évaluateurs d'identifier leur(s) option(s) préférée(s), mais la présentation des options peut être plus utile pour la direction. C'est une approche quelque peu similaire au fait de demander aux évaluateurs de présenter des conclusions et de développer ensuite des recommandations dans un atelier avec les parties prenantes comprenant les utilisateurs de l'évaluation.

## 6.3 Le nombre de questions

Souvent, les TdR d'une évaluation énoncent un trop grand nombre de questions. Même si l'on trouve quelques exemples où ce n'est pas le cas, comme l'évaluation de l'impact de l'aide alimentaire sur les réfugiés de longue durée (Sutter *et al.*, 2012) qui ne comportaient que quatre questions d'évaluation, il n'est pas rare que des TdR en identifient vingt ou plus.

Dans une étude portant sur la qualité des évaluations du DFID, Manning a recommandé de limiter le nombre de questions d'évaluation « de manière à éviter que les rapports ne soient trop longs et discursifs » (2009: xi). Dans une étude connexe, Perrin notait que même les évaluations allégées du DFID demandent des réponses à des dizaines de questions et que le manque de « focalisation causait souvent une attention superficielle à de trop nombreux éléments et débouchait *in fine* sur une longue liste de conclusions où les conclusions et les implications les plus importantes sont difficiles à identifier » (2009: vii).

Le problème lié au fait de poser trop de questions d'évaluation est encore plus important pour l'EAH car les évaluations de l'action humanitaire sont souvent réalisées avec des budgets relativement modestes.

Voici plusieurs bonnes raisons de limiter le nombre de questions :

1. **Se focaliser sur l'évaluation.** Comme Perrin (2009: vii) le remarquait dans son étude sur les évaluations de DFID : « Quand tout est prioritaire..., plus rien ne l'est ». USAID (1996: 2) notait également : « Poser trop de questions peut donner lieu à un effort désorganisé ». Le nombre de questions d'évaluation doit par conséquent se limiter aux domaines d'intérêt prioritaires. S'il y a trop de questions, l'équipe d'évaluation peut donner la priorité à celles qui intéressent le plus, ce qui peut ne pas coïncider avec les priorités de l'organisation commanditaire.
2. **Pour s'assurer que l'équipe d'évaluation traite les questions suffisamment en profondeur.** S'il y a 25 questions, elle ne pourra consacrer qu'un temps très limité à traiter chacune d'elles. Cela aboutit à des rapports d'évaluation superficiels.
3. **Pour rendre l'évaluation plus utile.** Quelques organisations peuvent gérer plus d'une demi-douzaine de recommandations à un niveau particulier. Chaque question d'évaluation demande une réponse, et les réponses sont généralement transcrites en résultats, lesquelles donnent lieu à des conclusions. Même si une simple recommandation peut s'appuyer sur plusieurs conclusions, il est plus courant qu'une conclusion amène plusieurs recommandations.  
Trop de recommandations compliqueront l'utilisation de l'évaluation.

Les départements des évaluations commencent progressivement à limiter le nombre de questions d'évaluation de manière à traiter ces points (Riddle, 2009: 16).

Hallam et Bonino (2013) notent que :

Le CICR demandait par le passé à toutes les parties intéressées d'identifier la gamme de questions et de thématiques qu'elles aimeraient voir aborder dans une évaluation. Le département des évaluations les transformait ensuite en questions « évaluables ». Toutefois, le périmètre des évaluations ne cessait d'augmenter jusqu'à devenir ingérable.

Pour réduire ce problème, le département des évaluations essaie désormais de se focaliser sur seulement trois questions clés pour chaque évaluation. (p. 62)

Le logiciel utilisé par NRC pour générer les TdR d'évaluations n'autorise que trois questions d'évaluation. Method Labs a créé une liste de questions directrices permettant de réduire le périmètre de la collecte de données (Buffardi, 2015).

## 6.4 Décomposer les grandes questions

Comment faites-vous pour évaluer des questions de premier niveau comme : « Avons-nous adopté la bonne stratégie ? » Il est difficile de répondre à de telles questions de façon rigoureuse.

Pour les décomposer, une méthode consiste à envisager la manière dont la réponse à la question de premier niveau pourrait être utilisée. Il est clair que la raison pour laquelle on pose une telle question est de décider si l'on doit conserver la stratégie en cours ou l'améliorer. Il peut également être intéressant de comparer des stratégies avec celles développées par d'autres acteurs.

Les questions sous-jacentes sont probablement les suivantes :

- Comment pouvons-nous améliorer notre stratégie actuelle ? (orientée vers l'action)
- Quels sont les avantages ou les inconvénients de notre stratégie actuelle par rapport à celles employées dans des contextes similaires ? (normative)

En résumé, observez ce que l'organisation commanditaire pourra faire avec la réponse à une question et voyez ensuite si cela peut être affiné jusqu'à fournir une meilleure base pour l'action. Même la première de ces questions sous-jacentes est probablement trop vaste. La stratégie en cours comprendra une combinaison d'éléments.

Ainsi, les questions peuvent être :

- Quels ajouts, suppressions ou modifications pourrions-nous faire pour augmenter l'impact ou réduire les coûts ? (orientée vers l'action)

Comment fonctionne notre stratégie par rapport à celles d'autres organisations et/ou contextes ? (normative)

Dans ce cas, il serait pertinent de traiter en premier la comparaison avec des stratégies choisies dans d'autres contextes puisque cela se prête à une [revue documentaire](#). Celle-ci peut déboucher sur une enquête afin d'examiner des éléments de la stratégie en cours.

Ce processus produit des questions auxquelles il est plus facile de répondre de manière rigoureuse. Il faut néanmoins répondre aux questions évaluatives sous-jacentes parce que la comparaison de stratégie ou d'alternatives implique des jugements évaluatifs, mais ces jugements sont désormais formulés dans le cadre d'actions pratiques. Voir [Section 16](#) : [L'analyse](#) pour des éléments relatifs à la manière de répondre aux questions évaluatives.

Dans certains cas, l'équipe d'évaluation doit décomposer les questions de premier niveau en questions d'évaluation auxquelles elle s'efforce de répondre, et les présente dans le [rapport de démarrage](#), souvent dans une matrice d'évaluation. Cependant, l'organisation commanditaire peut être mieux placée pour déterminer comment les réponses aux questions d'évaluation pourraient être utilisées.

## 6.5 Critères d'évaluation



### Astuce :

Les questions d'évaluation viennent en premier.

Les critères d'évaluation et les autres cadres sont des outils utiles une fois que vous avez choisi vos questions d'évaluation. Critères et cadres sont de bons serveurs mais de mauvais maîtres. Evitez de lister vos critères avant de choisir vos questions pour chaque critère et restez bien focalisé sur la manière dont conclusions et recommandations [seront utilisées](#).

Nous sommes engagés dans des processus d'évaluation au quotidien. En achetant des pommes au supermarché ou en choisissant un sandwich à la cafétéria, nous émettons un jugement sur les options qui s'offrent à nous en utilisant des critères personnels de qualité et de valeur. Une EAH formelle utilise une approche bien plus structurée, mais se concentre globalement sur des critères similaires.

En 1991, l'OCDE/CAD, se focalisant sur les problèmes les plus communément rencontrés dans les projets de développement, a proposé quatre critères de qualité - pertinence, efficacité, durabilité et impact – et le critère de valeur de l'efficacité (OCDE/CAD, 1991). Quelques années plus tard, l'OCDE/CAD a adapté ces critères pour l'EAH dans des situations d'urgence complexes (OCDE/CAD, 1999), ajoutant les critères de couverture et de cohérence, mais suggérant également que le caractère approprié (« *appropriateness* ») constitue une alternative à la pertinence et l'interdépendance une alternative à la durabilité. En outre, l'organisation a proposé deux nouveaux critères : la coordination et la protection. Ces critères reflétaient les plus grands problèmes rencontrés par l'action humanitaire dans les années 90.

Malgré les défauts des critères OECD/CAD ne sont pas parfaits, il existe de bonnes raisons pour les utiliser afin de vérifier vos questions d'évaluation :

- Utiliser les critères standards rend beaucoup plus facile le processus de méta-évaluation (tirer des leçons d'un large éventail d'évaluations).
- Les critères standards sont susceptibles d'identifier les points faibles récurrents dans l'action humanitaire puisqu'ils s'appuient sur l'expérience et la recherche.
- Il est plus facile pour les responsables d'évaluations et les évaluateurs de réaliser des évaluations qui utilisent les critères standards.



**Astuce :**

Vous n'avez pas besoin de couvrir tous les critères. Vos questions d'évaluation peuvent en aborder seulement un ou deux.

Durant la phase de planification de l'EAH, identifiez tout d'abord ce qu'il vous faut savoir, puis classez ces points au sein de critères d'évaluation – et non le contraire. Les critères sont des outils de réflexion qui peuvent faire naître de nouvelles questions pertinentes. Posez seulement des questions d'évaluation sur lesquelles vous êtes prêts à agir en vous appuyant sur les réponses. N'utilisez que les critères qui concernent les questions auxquelles vous voulez répondre. Ce sont les questions qui comptent, pas les critères.

Une seule question d'évaluation peut recouvrir plusieurs critères. Par exemple, la question « Dans quelle mesure notre stratégie de ciblage nous a-t-elle permis de fournir une assistance au bon moment et au meilleur coût ? » englobe les critères de couverture (ciblage), d'efficacité (moment opportun) et d'efficience (rentabilité). Vous pouvez encore décomposer cette question, mais toutes les questions relatives à l'efficacité ou à l'efficience du ciblage incluront inévitablement le critère de la couverture.

Dans le [Tableau 6.2](#), le principal critère est identifié pour chacun des exemples de question, avec des commentaires et une indication de l'action qu'une organisation pourrait entreprendre en réaction aux réponses. De plus, une question alternative focalisée sur l'utilisation est également donnée, ce qui peut aider à rendre les rapports d'évaluation moins discursifs.



**Astuce :**

Utilisez des questions orientées vers l'action pour les évaluations de plus petite taille. Ces questions se concentrent directement sur ce qu'une organisation pourrait faire pour améliorer ses performances à l'avenir et sont donc particulièrement appropriées pour les petites évaluations.

**Pour approfondir : Questions et critères d'évaluation**

Tableau 6.2 : Questions et critères d'évaluation

Exemples de question	Critère principal	Définition du critère
<p>Dans quelle mesure notre programme a-t-il satisfait les besoins immédiats ?</p> <p>Comment pouvons-nous garantir que nos kits de cuisine sont plus adaptés aux besoins locaux ?</p>	Caractère approprié (" <i>appropriateness</i> ")	Proportion selon laquelle les activités humanitaires sont adaptées aux besoins locaux, et augmentent en fonction l'appropriation, la redevabilité et la rentabilité. (Remplace le critère de pertinence utilisé dans les évaluations de développement)
<p>Dans quelle mesure le programme a-t-il atteint ses objectifs, notamment la livraison à temps de l'aide d'urgence ?</p> <p>Quels changements pouvons-nous apporter pour réduire le taux d'abandon dans les programmes d'alimentation complémentaire?</p>	Efficacité	Proportion selon laquelle une activité atteint son objectif, ou dans quelle mesure cela pourrait arriver étant donné les résultats.
<p>Notre programme abri a-t-il présenté un bon ratio coût-efficacité ?</p> <p>Comment pouvons-nous réduire les temps d'attente à la clinique ?</p>	Efficience	Les produits – qualitatifs et quantitatifs – réalisés comme résultat des moyens.
<p>Quel a été l'impact positif et négatif du programme de transferts monétaires et de bons ?</p> <p>Quelles mesures pourrait-on prendre pour réduire les dégâts causés par la collecte de bois de chauffe des réfugiés ?</p>	Impact	Les effets plus larges du projet – sociaux, économiques, techniques et environnementaux – sur les individus, les groupes des deux sexes et d'âge, les communautés et les institutions. Les impacts peuvent être intentionnels ou non, positifs et négatifs, macro (sectoriels) et micro (ménages). (Ce n'est pas exactement la même chose que l'« impact » dans la chaîne de résultats)



 **Tableau 6.2 : Questions et critères d'évaluation**

Exemples de question	Critère principal	Définition du critère
<p>Comment la fourniture de vaccins gratuits pour le bétail a-t-elle affecté la méthode de recouvrement des coûts des acteurs communautaires de la santé animale ?</p> <p>Que pouvons-nous faire pour empêcher que les distributions alimentaires ne découragent les agriculteurs de planter ?</p>	Interdépendance	Proportion selon laquelle les activités relevant de l'urgence à court terme sont réalisées dans un contexte qui prend en compte les problèmes à plus long terme et interdépendants. Remplace le critère de durabilité utilisé dans les évaluations de développement.
<p>Dans quelle mesure les transferts monétaires ont-ils bénéficié aux ménages les plus vulnérables ?</p> <p>Comment pouvons-nous garantir que les groupes et les individus marginalisés ont également accès aux subventions pour les abris?</p>	Couverture	Proportion selon laquelle les principaux groupes de population dont la vie est menacée ont été atteints par l'action humanitaire.
<p>Dans quelle mesure les politiques des organisations sont-elles cohérentes en matière de protection, et quelles en sont les implications ?</p> <p>Comment assurer un plaidoyer pour que d'autres bailleurs prennent en compte les droits de l'Homme dans leurs décisions de financement ?</p>	Cohérence	Proportion selon laquelle les politiques sécuritaires, développementales, commerciales et militaires ainsi que les politiques humanitaires sont cohérentes et prennent en compte les considérations humanitaires et relatives aux droits de l'homme. (Plus focalisé sur les politiques du bailleur, mais peut également s'appliquer à la cohérence des politiques d'organisations individuelles)
<p>Dans quelle mesure la distribution alimentaire a-t-elle été bien coordonnée, au niveau régional, et avec quelles conséquences ?</p> <p>Comment pouvons-nous réduire les lacunes relatives à l'approvisionnement en eau par les différentes organisations utilisant des citernes à eau ?</p>	Coordination	Proportion selon laquelle les interventions de différents acteurs sont en phase avec d'autres, promeuvent la synergie, évitent les lacunes, les duplications et les conflits de ressources. (Souvent regroupée avec l'efficacité)

Source : Adapté de Beck (2006), qui fournit de plus amples informations sur l'utilisation des critères CAD lors d'évaluations

Evitez de commencer par les critères et de choisir ensuite vos questions pour chacun d'eux, pour des raisons de limitations des ressources mais aussi parce qu'une EAH doit se focaliser sur les besoins des utilisateurs.



Les questions doivent orienter l'évaluation humanitaire, les critères d'évaluation servant seulement de liste de vérification pour s'assurer qu'aucun problème majeur n'a été laissé de côté. Perrin (2009: xi) recommandait que les évaluations de DFID « se focalisent sur un nombre limité de priorités et de sujets clés, utilisant les critères du CAD de façon intelligente plutôt que de façon mécanique, dans le but d'identifier un petit nombre de questions d'évaluation qui soient réalistes à traiter, qui prennent en compte le contexte, la disponibilité des données, ainsi que le périmètre et les ressources prévus pour l'évaluation ».

Pour s'assurer que les évaluations sont complètes et comparables, il se peut que certaines organisations veuillent montrer quels critères d'évaluation sont couverts par quelles questions spécifiques. Dans certains cas, une question peut aborder plusieurs critères, mais cela ne suffit pas toujours à couvrir tous les critères.

### Le problème de la circulation des termes de référence

Lorsque les termes de référence (TdR) circulent pour être commentés, le nombre de questions d'évaluation augmente car chaque département/partie prenante ajoute quelques suggestions. Aussi, plus la circulation est large, plus le nombre de questions est important. Hallam et Bonino (2013) font remarquer que les « évaluations de grande qualité qui traitent les vrais besoins en matière d'informations ... *augmentent à leur tour la demande d'évaluations* ». Les questions ajoutées lors de la circulation des TdR sont souvent mal prises en compte et plus susceptibles de diminuer la demande d'évaluation que de l'augmenter.

Parmi les stratégies permettant de contrôler le nombre de questions, on peut noter les suivantes :

- Inclure seulement les objectifs globaux de l'évaluation dans les TdR et laisser l'équipe développer les questions d'évaluation dans le rapport de démarrage.
- Demander aux relecteurs de classer les questions suggérées, puis de classer et justifier toute question complémentaire qu'ils proposent.
- Demander à ceux qui suggèrent des questions complémentaires d'expliquer comment ils utiliseraient les réponses et comment cela justifierait des coûts additionnels.



**Astuce :**

Faire circuler les TdR sous format PDF plutôt que dans un document facilement modifiable. Cela décourage les ajouts de questions à tort et à travers.

## 6.6 Choisir les meilleures questions

Si vous arrivez à un grand nombre de questions, vous devrez les classer. Pour ce faire, une solution consiste à les noter en fonction de différents facteurs, notamment dans quelle mesure les réponses à ces questions satisferont les besoins des utilisateurs. Demandez-vous par exemple si les questions :

- Sont centrales par rapport à l'objectif de l'évaluation
- Peuvent voir leurs réponses immédiatement mises en application
- Peuvent trouver une réponse sans équivoque dans le cadre de l'évaluation
- Peuvent trouver une réponse seulement au moyen d'une évaluation
- Sont centrales par rapport à votre mandat principal
- Peuvent améliorer la qualité du service que vous délivrez
- Peuvent réduire les coûts par personne du service délivré
- Affecteront ou seront utiles à presque toutes les parties prenantes.

Les facteurs, ou la pondération que vous leur donnez, varieront en fonction des buts de l'évaluation. Le résultat peut être présenté sous la forme d'une rubrique (voir Tableau 6.3). Il faut enfin décider à l'avance quels facteurs sont importants et noter les questions potentielles en fonction.

**Tableau 6.3 : Rubrique pour évaluer les questions d'évaluation**

Question par facteur Appliquez la question par facteur à votre potentielle question d'évaluation	Note					Pondération* Quelle est l'importance de la question par facteur pour votre évaluation ?	Note pour chaque question potentielle (note x pondération)							
	Note = 1	Note = 2	Note = 3	Note = 4	Note = 5		Question potentielle 1	Question potentielle 2	Question potentielle 3	Question potentielle 4	Question potentielle 5			
A Si nous avions la réponse à cette question, nous pourrions...?	Être mieux informé sur notre travail	L'utiliser pour plaider en faveur de changements dans notre manière de travailler	L'appliquer dans les futurs programmes	L'appliquer immédiatement dans nos programmes en cours										
B L'évaluation peut-elle répondre à cette question sans équivoque avec les ressources disponibles ?	Peu probable	Possible	Probable	Oui										
C Pourquoi ne peut-on répondre à cette question avec d'autres moyens ?	Oui, mais nous disposons d'un budget pour l'évaluation	Une évaluation plus pratique pour répondre à cette question	Une évaluation est la manière la moins coûteuse de répondre à cette question	Non, seule une évaluation peut répondre efficacement à cette question										
D Dans quelle mesure cette question est-elle centrale par rapport à votre mandat principal ?	Elle est périphérique par rapport à notre mandat	Elle est pertinente par rapport à notre mandat principal.	Nous pensons qu'elle fera partie principal à l'avenir.	Elle est centrale par rapport à notre mandat principal.										
E Quel est l'impact potentiel de la réponse à cette question sur la qualité des services que nous délivrons ?	Faible impact sur la qualité de nos services	Des impacts sur la qualité de nos services	Fort impact sur les services que nous délivrons	Elle peut améliorer de façon substantielle la qualité des services que nous délivrons.										
F Quel est l'impact potentiel de la réponse à cette question sur nos services délivrés ?	Faible impact sur le coût de nos services	Elle peut réduire certains coûts économiques	Elle peut réduire certaines dépenses économiques en termes de coût par personne	Elle peut réduire de façon significative le coût des services que nous délivrons.										
G Quelle proportion de notre groupe cible serait affectée par la réponse à cette question ?	Faible	Faible, mais ces groupes sont particulièrement vulnérables	Importante, notamment les groupes qui sont particulièrement vulnérables	Presque la totalité.										

\* Certains facteurs (A-C) doivent avoir plus d'importance dans le cadre de certaines évaluations et être pondérés en conséquence.

# 7 / Les termes de référence et la budgétisation d'une évaluation

Cette section couvre les termes de référence (TdR)<sup>3</sup> d'une évaluation. Les TdR sont représentés dans un document qui expose le travail d'évaluation de façon suffisamment détaillée pour que les évaluateurs potentiels comprennent ce qui est attendu.

Cette section couvre également de façon succincte la question du budget, à la fois pour que le responsable de l'évaluation comprenne ce qui est potentiellement faisable et pour que l'équipe d'évaluation prépare son offre.

## 7.1 Termes de référence (TdR)



### Définition : TdR

Les TdR présentent : « une vue d'ensemble des exigences et des attentes de l'évaluation. Ils fournissent un énoncé explicite des objectifs de l'évaluation, des rôles et des responsabilités des évaluateurs et du client de l'évaluation ainsi que des ressources disponibles pour l'évaluation » (Roberts *et al.*, 2011 : 2).

Les TdR forment la base du contrat entre l'équipe d'évaluation et l'organisation commanditaire.

Les TdR sont d'une importance critique pour établir la qualité de l'évaluation car les choix relatifs au périmètre et à l'approche peuvent avoir un grand impact sur la façon dont le résultat sera utilisable.

Les TdR définissent le périmètre, en termes d'éléments du projet ou du programme, le calendrier, et les interventions spécifiques les plus à même de vous fournir des réponses utilisables. Les TdR peuvent également exposer l'approche à utiliser pour l'évaluation et le budget permettant de répondre aux questions. Ils clarifient si l'évaluation obtiendra de meilleures réponses en étant centrée sur le personnel, les organisations partenaires ou les populations affectées.

## 7.1.1 Les TdR et le rapport de démarrage définissent l'évaluation

Davidson (2012) souligne le fait que l'évaluation est une entreprise conjointe entre l'organisation commanditaire et les évaluateurs.

Une évaluation de grande qualité, intéressante et « applicable » ne dépend pas que de la qualification technique de l'évaluateur et de ses compétences en matière de consultation. Les décisions prises et les actions engagées (ou non) par le client peuvent faire ou défaire la valeur de l'évaluation pour une organisation.

Une évaluation de grande valeur est le produit d'une interaction fructueuse entre un client bien informé et une équipe d'évaluation réactive dotée des compétences appropriées.

Le responsable de l'évaluation a deux options pour impliquer l'équipe d'évaluation dans la négociation du périmètre de l'évaluation.

Une approche considère les TdR comme un brouillon tant que l'équipe d'évaluation n'a pas été recrutée. Il est convenu que les questions en suspens font partie de la finalisation du contrat. C'est une bonne pratique, mais les systèmes financiers de certaines organisations écartent ce type de négociations.<sup>4</sup>

Une autre approche consiste à ce que l'équipe d'évaluation utilise le rapport de démarrage pour proposer la manière dont elle comprend les TdR. Le responsable de l'évaluation et l'équipe d'évaluation peuvent utiliser le rapport de démarrage comme base de négociation du périmètre final de l'évaluation. Une mission de démarrage peut contribuer à la qualité du périmètre final de l'évaluation.

La combinaison des TdR et du rapport de démarrage définit ce *qui* doit être évalué et *pourquoi*, ainsi que *comment* cela va être évalué. C'est pourquoi la liste de vérification d'UNEG couvre à la fois les TdR et le rapport de démarrage (UNEG, 2010b). Le responsable de l'évaluation peut soit définir certains éléments à l'avance, soit laisser à l'équipe d'évaluation le soin de les traiter dans le rapport de démarrage.

Tableau 7.1 : Ce qu'il faut avoir et ce qu'il est bon d'avoir dans les TdR et le rapport de démarrage d'une évaluation

<p><b>Éléments généralement inclus dans les TdR</b></p>	<ul style="list-style-type: none"> <li>• Contexte</li> <li>• Objectif général (comment l'évaluation sera utilisée)</li> <li>• Objectifs spécifiques</li> <li>• Critères</li> <li>• Périmètre</li> <li>• Public</li> <li>• Rôles et responsabilités</li> <li>• Etapes clés</li> <li>• Réalisations / produits attendus</li> <li>• Contenu du rapport de démarrage</li> </ul>
<p><b>Éléments qui peuvent être inclus dans les TdR ou le rapport de démarrage</b></p>	<ul style="list-style-type: none"> <li>• Cadre de l'évaluation</li> <li>• Questions de l'évaluation</li> <li>• Sources à utiliser</li> <li>• Matrice d'évaluation</li> <li>• Conceptions de l'évaluation</li> <li>• Méthodes de collecte des données</li> <li>• Indicateurs à mesurer</li> <li>• Méthodes d'analyse des données</li> <li>• Contenu du rapport d'évaluation</li> </ul>
<p><b>Éléments généralement inclus dans le rapport de démarrage</b></p>	<ul style="list-style-type: none"> <li>• Plan de travail</li> <li>• Allocation du travail au sein de l'équipe</li> <li>• Outils de collecte des données</li> </ul>

Les TdR explicitent ce que le responsable de l'évaluation attend de l'équipe d'évaluation, et le rapport de démarrage est la réponse de l'équipe aux TdR. Ainsi, les TdR et le rapport de démarrage forment un dialogue, complété par les échanges liés au rapport de démarrage. Les rapports de démarrage sont pertinents, y compris pour les évaluations de petite taille. La seule exception à cela est le cas où l'unité ou le département qui prépare les TdR réalisera également l'évaluation.

En règle générale, à moins qu'il y ait une raison impérieuse (comme le fait de s'attendre à ce que l'équipe d'évaluation soit relativement peu expérimentée), il est préférable de laisser autant que possible l'équipe expliquer son approche dans le rapport de démarrage plutôt que de la définir dans les TdR. Cela s'explique par le fait qu'au moment de la phase du rapport de démarrage, les sujets à évaluer doivent être déjà bien mieux identifiés. Il est également préférable de laisser à l'équipe d'évaluation le processus de présentation du produit visé.

Même pour des évaluations internes, des TdR doivent être préparés de manière à ce que l'objectif spécifique, l'objectif général et le périmètre soient clairement exposés. Cela évite que des incompréhensions ne surviennent plus tard.



## 7.1.2 Que trouve-t-on dans les TdR ?

Tous les TdR, y compris ceux des plus petites évaluations, comprennent les éléments suivants : (voir également [le tableau 7.1](#))

- Le contexte de l'évaluation
- Les objectifs de l'évaluation
- Le périmètre en termes de projet ou de programme couvert, les secteurs à évaluer, l'étendue géographique et la période prise en compte
- Les étapes clés, les dates butoir et les produits attendus
- Les rôles et les responsabilités de l'équipe d'évaluation et du responsable de l'évaluation
- Quelles ressources, y compris les données existantes, sont disponibles

Dans la plupart des cas, les TdR incluront également les questions d'évaluation auxquelles il faut répondre. Certains TdR énoncent seulement les objectifs de l'évaluation et laissent à l'équipe le soin de proposer des questions (soumises à la validation du responsable de l'évaluation) dans le [rapport de démarrage](#).

Parmi les guides de rédaction de TdR, on peut citer :

- Le modèle des TdR de DFID, document bref (2 pages) qui va à l'essentiel (DFID, 2010)
- La liste de vérification de la qualité de l'UNEG qui porte sur les TdR et les rapports de démarrage (UNEG, 2010)
- Le guide sur la préparation des TdR de la Banque mondiale (Roberts *et al.*, 2011) fournit des conseils détaillés
- Le guide d'USAID sur la rédaction d'un périmètre de travail (USAID, 1996)



### Les évaluations de petite taille peuvent inclure des questions contractuelles dans les TdR

Dans le cas de petites évaluations, les TdR peuvent également intégrer les détails relatifs à la procédure d'appels d'offre. Dans ce cas, les TdR indiqueront les dates butoir de l'offre, la documentation à soumettre avec l'offre, la manière dont les offres seront notées, et toute autre exigence spécifique (par exemple, présenter de façon distincte les offres techniques et financières). Des évaluations de grande taille peuvent exposer ces détails dans la demande de propositions et présenter seulement les détails de l'évaluation dans les TdR.

Tableau 7.2 : Éléments potentiels des termes de référence


Élément	Commentaires
Historique	Cela ne nécessite qu'un ou deux paragraphe(s) ; la plus grande partie des connaissances contextuelles de l'équipe d'évaluation viendra de lectures préliminaires. Les introductions de propositions de projet ou les documents d'appels lancés sont de bonnes sources.
Objectif de l'évaluation	L'évaluation vise-t-elle l'apprentissage ou la redevabilité ? S'il s'agit des deux, quel objectif est prioritaire ? Par exemple, une évaluation peut chercher à informer les bailleurs de l'efficacité avec laquelle leur argent a été utilisé (redevabilité) mais peut également examiner les facteurs qui, dans la conception et la mise en œuvre du projet, ont conduit aux projets les plus efficaces (apprentissage). Toutes les évaluations, qu'elles soient principalement orientées vers la redevabilité ou l'apprentissage, fournissent des opportunités d'apprentissage.
Contexte de l'évaluation	Pourquoi faire l'évaluation maintenant ? Y a-t-il des dates butoir internes ou externes ? L'évaluation est-elle liée à des décisions d'un cycle de financement ou à des décisions relatives à la future stratégie de l'organisation ? De telles dates butoir et imbrications doivent être claires si vous impliquez les principales parties prenantes dès le début, et doivent apparaître clairement dans les TdR. Gardez à l'esprit que cela est lié à l'utilisation, voir <a href="#">Section 3</a> .
Périmètre de l'évaluation	Quels secteurs, étendue géographique, phase d'intervention et période de temps l'évaluation va-t-elle couvrir ? Se concentrera-t-elle sur le niveau des politiques ? Le niveau du programme ? Des processus ou des opérations spécifiques ? S'agit-il d'une évaluation en temps réel, d'une évaluation à mi-parcours ou d'une évaluation <i>ex-post</i> ? S'agit-il d'une évaluation d'une seule organisation, d'une auto-évaluation ou d'une évaluation conjointe ? Cela est lié au type d'évaluation, voir <a href="#">Section 4</a> .



Élément	Commentaires
Utilisateurs	Comment s'attend-on à ce que l'évaluation soit utilisée ? La réponse à cette question déterminera la longueur et la lisibilité des produits de l'évaluation. Voir <a href="#">Section 3 : Penser tôt et souvent à l'utilisation de l'évaluation</a> .
Le cadre de l'évaluation	Existe-t-il un modèle conceptuel que vous souhaitez voir utilisé par les membres de l'équipe d'évaluation lors de l'élaboration de leur méthodologie et de leur analyse – par exemple, le cadre des moyens de subsistance relatif aux urgences humanitaires complexes ? Quelles sont les normes internationales pertinentes pour cette évaluation – par exemple, les standards SPHERE ou la Norme humanitaire fondamentale (CHS) ? Voir <a href="#">Section 5 : Elaborer votre évaluation</a> .
Principales questions d'évaluation	<p>Quelles sont les principales questions auxquelles vous voulez que l'évaluation réponde ? Quels critères OCDE/CAD voulez vous utiliser et comment sont ils reliés ? Voir <a href="#">Section 6</a> sur les questions et les critères OCDE/CAD.</p> <div data-bbox="470 730 1083 847" style="background-color: #f0f0f0; padding: 5px; border: 1px solid #ccc;"> <p> <b>Astuce :</b> Peut-être préférez-vous simplement énoncer les objectifs et laisser l'équipe d'évaluation proposer des questions dans le rapport de démarrage.</p> </div> <div data-bbox="470 863 1083 975" style="background-color: #f0f0f0; padding: 5px; border: 1px solid #ccc;"> <p> <b>Astuce :</b> Conservez un faible nombre de questions d'évaluation afin que l'évaluation reste ciblée.</p> </div>
Phase de démarrage	<p>Les TdR doivent clarifier quelles activités sont attendues dans la phase de démarrage. Celle-ci se limite-t-elle à de la recherche documentaire ou comprend-elle une visite de démarrage ?</p> <p>Les TdR doivent également exposer le contenu attendu du rapport de démarrage. Voir <a href="#">Section 8 : La phase de démarrage</a>.</p>
Conceptions	Une évaluation peut utiliser différentes conceptions. Un type de question donné peut obtenir de meilleures réponses avec une approche particulière, mais les contraintes logistiques et budgétaires peuvent limiter les choix. En général, l'équipe propose des approches dans le rapport de démarrage, mais l'organisation commanditaire peut spécifier qu'une approche particulière doit être utilisée – par exemple, pour être conforme à la politique générale de l'organisation. Voir <a href="#">Section 11 : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation</a> .



← Tableau 7.2 : Éléments potentiels des termes de référence

Élément	Commentaires
Méthodes de collecte des données	Cela ne doit pas faire l'objet d'une description détaillée, mais plutôt d'une indication des préférences méthodologiques – par exemple, si vous souhaitez que l'équipe consulte des personnes en particulier, comme des représentants du gouvernement ou la population affectée, ou bien qu'elle utilise des méthodes particulières, comme une enquête officielle. En général, l'équipe doit développer la méthodologie détaillée durant la phase de démarrage. Si le responsable de l'évaluation spécifie à la fois le produit et le processus, l'équipe d'évaluation peut éprouver un moins grand sentiment d'appropriation et de responsabilité vis-à-vis du résultat final. Voir <a href="#">Section 10 : Les méthodes de recherche documentaire</a> et la <a href="#">Section 13 : Les méthodes de terrain</a> .
Indicateurs	C'est généralement l'équipe qui propose des indicateurs (si des indicateurs quantitatifs sont utilisés) dans le rapport de démarrage, mais l'organisation commanditaire peut demander l'utilisation d'indicateurs particuliers (par exemple, permettre une comparaison ou une synthèse avec d'autres évaluations).
Méthodes d'analyse des données	Les TdR doivent préciser toute exigence relative à l'utilisation d'une approche analytique en particulier. En général, l'équipe propose son approche analytique dans le rapport de démarrage. Voir <a href="#">Section 16 : L'analyse</a> .
Calendrier	<p>Spécifiez des dates pour les livrables clés et toute autre date butoir. Demandez aux parties prenantes si ces dates posent problème – par exemple, si elles tombent pendant la saison des pluies, la saison des récoltes, ou des périodes administrativement chargées. Voir <a href="#">Section 9 : Planifier et gérer votre évaluation</a> pour plus d'éléments sur les</p> <div data-bbox="494 1034 1107 1193" style="background-color: #f0f0f0; padding: 10px; border: 1px solid #ccc;">  <p><b>Astuce :</b> Prévoyez suffisamment de temps pour faire circuler une version provisoire du rapport (en général 14 à 21 jours) et réviser le rapport (également deux ou trois semaines).</p> </div>
Rôles et responsabilités	Précisez qui est responsable du transport sur le terrain et de prendre les rendez-vous, etc. Vérifiez avec vos collègues sur le terrain qu'ils peuvent correctement soutenir l'équipe d'évaluation.



← Tableau 7.2 : Éléments potentiels des termes de référence

Élément	Commentaires
Modalités de gestion	Spécifiez s'il y a aura des groupes consultatifs, de pilotage ou de référence ainsi que leur composition et leurs rôles. Voir <a href="#">Section 9 : Planifier et gérer votre évaluation</a> .
Compétences et qualifications	De quelles compétences et qualifications l'équipe d'évaluation a-t-elle besoin pour mener à bien cette évaluation ? Cela peut comprendre une connaissance plus large du contexte (du pays, du secteur, ou de l'organisation), des langues, ou des compétences dans des méthodes ou des formes de communication particulières. Voir <a href="#">Section 9 : Planifier et gérer votre évaluation pour plus d'éléments sur les compétences dont un évaluateur peut avoir besoin</a> .
Produits	<p>Les produits comprennent en général un rapport de démarrage, des comptes-rendus de visites sur le terrain (sous forme de notes ou de présentation), un rapport principal, un résumé de l'évaluation et des comptes-rendus des réunions ou de tout autre exercice considéré comme utile pour la diffusion des résultats. Précisez la longueur et le format si vous avez une préférence. Il est également utile de mettre en place des paiements par phase pour chaque produit spécifique afin d'encourager l'équipe d'évaluation à maintenir une approche focalisée sur son but. Voir <a href="#">Section 17 : Rédiger et communiquer des résultats d'évaluation en se focalisant sur l'utilisation</a>.</p> <p>Si le rapport de l'évaluation doit adhérer à un certain « style » propre à votre organisation, ou s'il existe de fortes préférences au niveau du format et du style, spécifiez-le afin d'éviter d'avoir à reprendre le rapport plus tard. Précisez la longueur du rapport (le nombre de mots est une mesure plus précise et moins susceptible de faire l'objet de malentendus que le nombre de pages) et quelles sont les annexes à inclure, si nécessaire. Vous pouvez également préciser la structure globale du rapport même s'il est préférable de donner à l'équipe d'évaluation une certaine flexibilité à ce sujet.</p>
Gestion des risques	Décrivez les risques et les défis susceptibles d'apparaître durant l'évaluation, et demandez à l'équipe comment elle envisage de les gérer. Voir <a href="#">Section 15 : L'accès restreint sur l'accès distant</a> .



3 / Penser tôt et souvent à l'utilisation de l'évaluation  
62 - 77

4 / Les types d'évaluation  
78 - 90

5 / Elaborer votre évaluation  
91 - 103

6 / Choisir les questions d'évaluation  
104 - 117

7 / Les termes de référence et la budgétisation d'une évaluation  
118 - 131

← Tableau 7.2 : Éléments potentiels des termes de référence

Élément	Commentaires
Budget	Donnez un budget indicatif pour l'évaluation. Voir <a href="#">Section 7 : Les termes de référence et la budgétisation d'une évaluation</a> .
Données disponibles	Identifier les principales sources de données documentaires qui seront disponibles pour l'équipe d'évaluation. Auront-ils accès à l'intranet, aux historiques des e-mails et à de la littérature « grise » (par exemple, les précédentes évaluations du projet ou programme)? Patricia Rogers remarque que l'une des causes d'échec des évaluations est le fait de « ne pas communiquer certaines données, de ne pas donner l'information que certaines données existent, de ne pas transmettre les évaluations précédentes, les évaluations concurrentes, les changements de politiques prévus, les changements de personnel à venir, les dates de décisions et de réunions cruciales » (commentaires envoyés à Davidson, 2012b).
Évaluation des offres	S'il n'y a pas un document d'appel d'offre formel, vous pouvez inclure les critères sur lesquels vous vous basez pour évaluer les différentes offres (quel pourcentage de la note ira au prix, à la composition de l'équipe, aux méthodes de collecte de données et d'analyse, et à d'autres critères) ainsi que la date butoir de réception des offres. Voir <a href="#">Section 9 : Planifier et gérer votre évaluation</a> .
Références clés	Vous pouvez joindre une liste des références clés pour l'évaluation. La plupart sont des documents internes, mais vérifiez auprès des principales parties prenantes si d'autres documents doivent être ajoutés aux lectures de démarrage. Voir <a href="#">Section 10 : Les méthodes de recherche documentaire</a> .

### 7.1.3 Le calendrier des TdR

Développer des TdR peut aussi bien prendre moins d'une semaine dans une petite organisation que plus d'un an dans une grande. Il est en général plus long de finaliser des TdR lorsque différentes parties prenantes se trouvent dans plusieurs pays.

Le IASC a développé une version provisoire convenue au préalable des TdR relatifs aux évaluations en temps réel inter-organisations des urgences humanitaires de manière à éviter de devoir négocier avec les parties prenantes à chaque fois (IASC, 2010). Ces TdR provisoires étaient élaborés avec l'idée que les contrats seraient signés avec l'équipe sélectionnée trente jours après la publication des TdR et des demandes de manifestation d'intérêt. Voir la [Section 9 : Planifier et gérer votre évaluation](#) pour plus d'éléments sur les calendriers d'une évaluation.

Certaines organisations prennent généralement plus de temps que le IASC. Une étude portant sur huit EAH auxquelles les auteurs ont participé montrait que le temps compris entre la publication des TdR et la signature du contrat allait de deux semaines à trois mois. La plupart des EAH ont pris deux mois entre la publication des TdR et la signature du contrat avec l'équipe. Un mois est probablement la durée moyenne qui convient pour les petites évaluations.

### 7.1.4 La longueur des TdR

Les TdR n'ont pas de longueur standard. Une étude portant sur un échantillonnage de convenance de trente EAH a montré que la plupart faisaient de deux à six pages et que trois faisaient plus de quinze pages. Les plus courts concernaient des évaluations recrutant un seul expert externe ou une équipe sélectionnée en interne.

Les évaluations plus complexes avaient des TdR plus détaillées.

**Astuce :**

Les TdR doivent refléter la complexité de la tâche. Pour une petite évaluation, la courte durée de la mission doit aller de pair avec des TdR plus brefs.

Le plus grand coût unique de toute évaluation employant des évaluateurs externes est le nombre de jours des consultants. Il correspond en général au produit de la taille de l'équipe et du nombre de jours de terrain multiplié par les honoraires et les taux de perdiems. Estimer le nombre total de jours constitue la première étape de l'estimation du coût de l'évaluation. Un budget peut être décomposé en phases de préparation, de terrain et de suivi.

## 7.2 Budgéter une évaluation

### La phase de démarrage

La phase préparatoire peut comprendre une première réunion pour discuter de l'évaluation, des lectures de fond et de la recherche documentaire, un entretien préliminaire et la rédaction d'un rapport de démarrage.

- La réunion initiale peut prendre un ou deux jours, en comptant les déplacements, mais tous les membres de l'équipe ne sont pas obligés d'y participer.
- Les lectures de fond, les entretiens préliminaires et le rapport initial peuvent prendre de 5 à 10 jours pour une petite évaluation, et plusieurs mois ou plus pour une évaluation plus vaste et plus complexe.

Une petite évaluation (une seule organisation avec une seule activité sur un seul site) prendra en général de trois à sept jours par évaluateur pour le travail préparatoire, en fonction de la nature de l'évaluation. Comme la rédaction d'un rapport, une préparation digne de ce nom prend souvent plus de temps que ce qui est budgété. Il est important de laisser une certaine flexibilité.

### Le travail de terrain

Les coûts liés au travail de terrain – avant tout le salaire du personnel et les dépenses en lien avec l'évaluation (par exemple, les billets d'avions, le transport local et l'interprétariat) – sont déterminés par la longueur du travail de terrain, ainsi que par les honoraires et les perdiems des évaluateurs (personnel et/ou externes).

- En règle générale, les méthodes quantitatives demandent davantage de travail de terrain et de plus larges équipes dotées de compétences techniques en conception et analyse statistique ; les méthodes qualitatives demandent un travail de terrain impliquant des compétences plus élevées.
- Une petite évaluation prend généralement une semaine d'observation et d'entretiens sur les sites choisis, avec une demi-semaine pour les entretiens du briefing initial et l'autre demi-semaine pour les entretiens de suivi et le débriefing. Il faut plus de temps s'il y a plusieurs lieux à visiter sur le terrain, ou si l'équipe est censée produire un rapport provisoire avant de quitter le pays.
- Une petite évaluation comprend généralement entre 12 et 14 jours de travail de terrain, transport compris ; une grande évaluation peut prendre des mois de travail de terrain.
- En règle générale, on compte un minimum d'une semaine de travail pour chaque pays visité, et une semaine de plus pour chaque site majeur (une province ou un district) à visiter.
- Si l'équipe s'est engagée dans un processus détaillé de consultation des populations affectées, comptez jusqu'à trois semaines par site.
- Les méthodes d'enquêtes quantitatives doivent être budgétées en fonction de la taille de l'échantillon et selon le temps nécessaire au traitement et à l'analyse des données.



**Tableau 7.3** : Estimer le nombre de jours des chefs d'équipe (CE) et des membres de l'équipe (ME)

Activité	Jours des CE	Jours des ME
Réunion de démarrage (comprenant le voyage)		
Entretiens préliminaires		
Recherche documentaire		
Visite de terrain préliminaire		
Développer les outils de collecte de données		
Rédiger le plan d'analyse		
Matrice d'évaluation		
Rapport de démarrage		
Réunion sur le rapport de démarrage		
Révision du rapport de démarrage		
Voyage jusqu'au terrain		
Briefing initial		
Réunions en capitale		
Travail de terrain 1		
Réunion de bilan à mi-parcours		
Travail de terrain 2		
Préparation du débriefing		
Débriefing		
Voyage de retour		
Collecte des données		
Analyse des données		
Première version du rapport d'évaluation		
Présentation du rapport		
Révision de la première version		
Révision de la seconde version		
Révision finale		

## Le travail de retour du terrain

- Cette phase comprend principalement l'assemblage et l'analyse des données collectées puis l'écriture des rapports. Encore une fois, elle est couramment sous-budgétée. Le fait d'assembler et d'analyser des données puis de rédiger le rapport demande presque toujours plus de temps que celui alloué, là encore en fonction de la complexité du rapport et de la quantité d'analyse requise.
- Les débriefings ont tendance à durer un à deux jours, mais le temps de voyage peut s'y ajouter. La rédaction du rapport peut prendre entre cinq et vingt jours, et chaque cycle de révision de deux à dix jours.
- Si une équipe est composée de nombreux membres, une bonne règle consiste à ajouter cinq jours à la préparation du rapport pour permettre au chef d'équipe d'incorporer leur contribution.
- Une petite évaluation consacre généralement entre sept et douze jours aux activités de retour du terrain.



### **Astuce :**

N'oubliez pas les coûts d'impression et de diffusion de votre rapport. N'attendez pas que la version finale soit prête avant de vous rendre compte qu'aucune ressource n'a été prévue pour la relecture ou la mise en page (images, infographie, etc.). La diffusion implique elle aussi de réels coûts (envoi d'intervenants à des réunions sectorielles ou régionales, etc.).

Les éléments courants de budget sont présentés dans le Tableau 7.4.

Tableau 7.4 : Eléments de coûts pour une évaluation

Ligne budgétaire	Éléments possibles
Personnel	Rémunérations et indemnités du personnel, indemnités du personnel des partenaires et autre personnel.
Consultants	Chef d'équipe, consultants internationaux et/ou nationaux, autres consultants.
Personnel de soutien	Rémunérations et indemnités de l'administration, des chercheurs travaillant sur le contexte, des interprètes, des chauffeurs, du personnel de sécurité, et autres.
Voyage	Visas, vols pour l'équipe d'évaluation et le personnel accompagnateur, vols pour les briefings et débriefings, transport local de l'équipe et du personnel accompagnateur.
Indemnités journalières	Coût du logement et du perdiem des consultants.
Saisie de données	Saisie et nettoyage de données pour supprimer les réponses qui n'ont pas de sens, comme quelqu'un enregistré à la fois comme un homme et enceinte.
Réunions et ateliers	Location du lieu, repas, et indemnités.
Production du rapport	Relecture, traduction, illustrations, infographie, mise en page, impression, développement de supports électroniques, distribution.
Autres produits	Coûts d'écriture, relecture, illustrations, etc., pour les documents de briefings, infographie, documents du type « 10 choses à savoir sur... »
Coûts de diffusion	Coûts de transport et honoraires pour la présentation des résultats au niveau national ou régional, ou à des réunions sectorielles, ou encore dans d'autres événements
Divers	Communication, e-mails et courriers postaux, téléconférences, licences et frais juridiques, sécurité.
Coûts administratifs	Souvent calculés sous la forme d'un pourcentage des autres coûts.

## Notes de fin de texte

### 4 / Les types d'évaluation

1. L'OCDE-CAD a développé un guide sur la gestion des évaluations conjointes (OCDE-CAD, 2006). Les évaluations conjointes de l'action humanitaire sont l'objet d'un chapitre de la septième revue de l'action humanitaire d'ALNAP (Beck et Buchanan-Smith, 2008).
2. Voir Breier, 2005.

### 7 / Les termes de référence et la budgétisation d'une évaluation

3. USAID appelle les TdR « périmètre de travail », ce qui est probablement une description plus précise.
4. Par exemple, le modèle de TdR de DIFD énonce que « les consultants ne seront pas impliqués en général » dans la rédaction du document (DFID, 2010:1).

3 / Penser tôt et souvent à l'utilisation de l'évaluation  
62 - 77

4 / Les types d'évaluation  
78 - 90

5 / Elaborer votre évaluation  
91 - 103

6 / Choisir les questions d'évaluation  
104 - 117

7 / Les termes de référence et la budgétisation d'une évaluation  
118 - 131

## Notes

# Planifier et concevoir l'évaluation



# 8 / La phase de démarrage

Cette section couvre la phase de démarrage de l'évaluation et, en particulier, le rapport de démarrage.



## Définition : Phase de démarrage

La phase de démarrage de l'évaluation va de la sélection de l'équipe d'évaluation à l'acceptation du rapport de démarrage.

Durant la phase de démarrage, l'équipe essaie d'appréhender pleinement le travail de l'évaluation et prépare un rapport exposant le plan de l'évaluation. La phase de démarrage a lieu avant le travail de terrain : son produit est le rapport de démarrage. Ce dernier peut être relativement court, comme dans le cas de l'évaluation du Fonds de réponse humanitaire aux situations d'urgence en Haïti (Morinière, 2011a), et peut être comparé avec le rapport d'évaluation final (Morinière, 2011b).

La plupart des TdR des évaluations humanitaires sont fixés plutôt que négociés entre le responsable de l'évaluation et l'équipe d'évaluation. Le rapport de démarrage permet à l'équipe d'évaluation de définir un plan spécifique et de trouver un accord à ce sujet avec le responsable de l'évaluation, mais aussi de soulever des points problématiques et de répondre à toute éventuelle ambiguïté des TdR. Par exemple, le travail sous-entendu dans les TdR peut prendre beaucoup plus de temps que ce qui est budgété. Le rapport de démarrage constitue donc une opportunité pour que l'équipe d'évaluation clarifie ce qu'elle peut et ne peut pas couvrir.

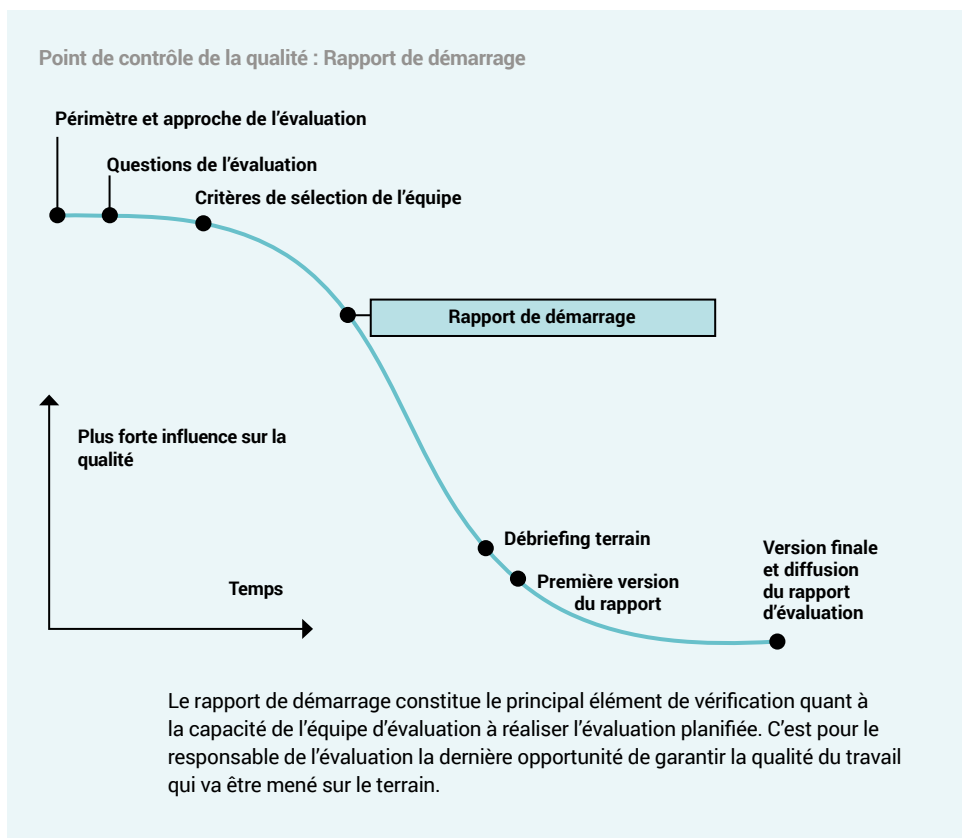


## Astuce :

Rédigez un rapport de démarrage, même pour les plus petites évaluations. Il est presque toujours utile que l'équipe d'évaluation produise un rapport de démarrage même si cela n'est pas demandé, car cela aide l'équipe à planifier son travail et à minimiser de possibles malentendus entre l'équipe et le responsable de l'évaluation.

## 8.1 Le rapport de démarrage et le contrôle qualité

Le rapport de démarrage est probablement l'outil le plus utile de la boîte à outils du contrôle qualité. De même qu'il aide à établir une compréhension commune du travail d'évaluation, il peut également indiquer au responsable de l'évaluation si l'équipe d'évaluation n'est pas en mesure de produire une évaluation de qualité. Toutefois, même si un rapport de démarrage de mauvaise qualité tend à confirmer que l'équipe d'évaluation n'est pas à la hauteur de la tâche, un rapport de bonne qualité ne garantit pas qu'elle le soit.





## L'essor du rapport de démarrage

Les rapports de démarrage et les matrices d'évaluation associées sont relativement récents dans l'EAH.

Un rapport de démarrage :

- Aide à ce que tout le monde commence à travailler avec la même base
- Aide à repérer des périmètres d'évaluation trop ambitieux
- Aide le responsable de l'évaluation à analyser comment l'équipe comprend l'évaluation et prévoit de l'aborder
- Fournit à l'équipe d'évaluation l'opportunité de convertir les TdR en un plan réaliste validé par le responsable de l'évaluation
- Donne à l'équipe d'évaluation une chance de clarifier les TdR et de souligner les tensions qui doivent être résolues par l'organisation commanditaire (par exemple, les attentes contradictoires au sein de l'organisation vis-à-vis de l'évaluation)
- Met en lumière les défauts de la conception et des méthodes proposées
- Donne aux autres parties prenantes une chance de recevoir une déclaration d'intention claire de la part de l'équipe d'évaluation de manière à ce qu'elles puissent rapidement repérer des difficultés potentielles liées à la méthodologie proposée
- Aide à identifier les équipes qui ne sont peut-être pas à la hauteur de la tâche
- Permet à l'équipe d'évaluation d'exposer clairement ce qu'elle va faire et, ce qui dans certains cas est plus important encore, ce qu'elle ne fera pas



### Astuce :

Pour une évaluation complexe, séparez les phases de démarrage et de travail de terrain. Il peut également être utile de réaliser des contrats distincts pour les phases de démarrage et de travail de terrain. Les candidats sont invités à faire des propositions pour les deux phases, mais le contrat relatif au travail de terrain est conditionné par l'acceptation du rapport de démarrage.

## 8.2 Les activités de la phase de démarrage

Le rapport de démarrage doit présenter un plan de travail et un calendrier clairs et réalistes qui prennent en compte les contraintes existantes. Le plan de travail doit inclure la répartition des rôles et des responsabilités au sein de l'équipe, ainsi que les dates butoir pour la remise des rapports internes et les plans de voyage détaillés. L'effort nécessaire pour mettre en place un plan d'évaluation clair dépend du périmètre et de la complexité de l'évaluation.

La phase de démarrage peut comprendre :

- Des entretiens préliminaires avec des informateurs clés pour appréhender le contexte
- Une revue documentaire, notamment une revue de la littérature et l'étude de données quantitatives secondaires
- Un atelier pour présenter une version provisoire du rapport de démarrage, ce qui peut être utile pour valider l'approche
- Développer et, si nécessaire, tester les outils de collecte des données
- Une visite préliminaire de cadrage dans le pays du travail de terrain, ce qui constitue une bonne pratique pour les évaluations complexes et de grande ampleur.



### Astuce :

Accordez-vous assez de temps. Laissez non seulement le temps nécessaire à ce que l'équipe réalise une revue documentaire et des entretiens préliminaires, mais aussi entre la remise du rapport de démarrage et le travail de terrain afin de pouvoir réaliser des ajustements.

## 8.3 La matrice d'évaluation

La plupart des rapports de démarrage comportent une matrice d'évaluation qui expose un plan de réponse à chacune des questions d'évaluation. Vous trouverez ci-dessous deux exemples, l'un provenant d'une évaluation d'un bailleur important et l'autre d'une évaluation d'une ONG de petite taille. Parce que les matrices d'évaluation sont relativement récentes, il n'existe pas de modèle convenu et les titres des colonnes peuvent varier.

Dans leur forme la plus simple, les matrices d'évaluation peuvent avoir les méthodes potentielles comme titres de colonnes et les questions comme titres de lignes. Les intersections seront mises en évidence d'une manière ou d'une autre (Todd *et al.*, 2015).

Ce Guide plaide pour l'utilisation d'une matrice à cinq colonnes :

- Questions de l'évaluation
- Critères d'évaluation (rappelez-vous qu'une seule question peut renvoyer à plus d'un critère)
- La conception à utiliser pour répondre à la question
- La méthode ou les méthodes à utiliser pour répondre à la question
- L'approche d'échantillonnage à utiliser pour la question.



#### Astuce :

Les responsables d'évaluation peuvent trouver utile de préparer leur propre matrice d'évaluation. Même si la préparation de la matrice d'évaluation est de la responsabilité de l'équipe d'évaluation, il peut être utile que le responsable d'évaluation en prépare une copie interne. Cette matrice ne doit pas être partagée avec l'équipe qui remporte l'appel d'offre tant que celle-ci n'a pas préparé sa propre version. Le responsable peut comparer les deux versions pour voir si cela révèle des lacunes majeures.

**Tableau 8.1** : Exemple d'une matrice d'évaluation utilisée pour un projet de transfert monétaire d'une petite ONG

Question	Critère	Conceptions	Méthodes	Echantillonnage
<b>Quel impact a eu le transfert monétaire sur la sécurité alimentaire des ménages ?</b>	Impact, efficacité	Différence de conception : comparer dans la durée les changements des scores des ménages en matière de sécurité alimentaire entre ménages bénéficiaires et non bénéficiaires	Enquête sur la sécurité alimentaire des ménages	Echantillonnage aléatoire utilisant le recensement du diagnostic initial de sécurité alimentaire comme cadre d'échantillonnage
		Non-expérimentale	Entretiens qualitatifs avec des membres des ménages sélectionnés	Sélection raisonnée de ménages présentant les scores les plus élevés et les plus bas dans le diagnostic initial

(Voir la [Section 11](#) : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation pour une description de la différence en matière de conception.

Veillez noter qu'une seule question peut impliquer plus d'une conception ou méthode, et que les méthodes individuelles peuvent aider à répondre à plus d'une question – non représenté ici.

Une fois la matrice d'évaluation complétée, elle peut être utilisée comme base d'un tableau de synthèse des observations, lequel peut faciliter l'écriture du rapport en rassemblant tous les éléments de preuve relatifs à une question ou une thématique pour un même lieu (pour plus d'informations sur les tableaux de synthèse des observations, voir [Section 16](#)).



#### **A garder à l'esprit**

Le processus de planification est toujours plus important que le plan, et l'équipe doit pouvoir s'approprier à sa façon les questions d'évaluation.

### **D'autres formats de matrices d'évaluation**

De nombreuses évaluations utilisent des matrices d'évaluation qui ont seulement les trois colonnes suivantes :

1. **Les questions auxquelles il faut répondre** : celles-ci peuvent être présentées sous la forme de questions clés avec des questions d'évaluation subsidiaires.
2. **Comment le jugement sera étayé** : les critères ou indicateurs à partir desquels l'équipe d'évaluation répondra à la question.
3. **Les sources et méthodes d'information attendues** : on retrouve généralement plusieurs sources et méthodes pour chacune des questions d'évaluation. Parfois, elles sont scindées en deux colonnes.

Les exemples suivants sont tirés des matrices de deux évaluations, l'une d'une évaluation multi-pays d'un bailleur, et l'autre de l'évaluation d'un seul projet. Les deux matrices ont quatre colonnes.

Une matrice d'évaluation de ce type montre comment l'équipe prévoit de répondre à chaque question et son analyse permet au responsable d'évaluation de voir si l'équipe a négligé toute source majeure.

Certains modèles de matrice sont plus complexes. Une matrice d'évaluation complétée peut être utilisée comme base d'un tableau de synthèse des observations, lequel rassemble tous les éléments de preuve relatifs à une question ou une thématique particulière pour un même lieu et peut être utilisé pour écrire le rapport (voir [Section 16](#)). Le [tableau 8.4](#) montre les grandes rubriques des matrices de trois évaluations inter-organisations.

Tableau 8.2 : Exemple de matrice d'évaluation utilisée pour une grande évaluation

Les grandes questions d'évaluation / sous-questions	Indicateurs	Méthodes analytiques	Sources des données
<p><b>1. Dans quelle mesure la stratégie humanitaire danoise est-elle pertinente et flexible étant donné le contexte humanitaire changeant depuis 2010 ?</b></p> <p><b>1.1 Les priorités stratégiques ont-elles été pertinentes étant donné l'évolution des défis humanitaires ?</b></p>	<p><b>1.1a</b> Nombre de priorités stratégiques couvertes par les programmes financées par Danida</p> <p><b>1.1b</b> Correspondance entre les priorités stratégiques et ce que Danida et ses partenaires considèrent comme des défis humanitaires clés</p> <p><b>1.1c</b> Capacités d'anticipation, d'adaptation et d'innovation des partenaires en matière d'identification et de gestion des nouveaux types de menaces et les opportunités de les atténuer</p> <p><b>1.1d</b> Données probantes indiquant que les stratégies de financement et pays de Danida sont assez flexibles pour permettre aux partenaires de s'adapter aux contextes changeants.</p>	<p>Analyse du portfolio, suivi des résultats et analyse comparative des partenaires pour évaluer la couverture des priorités stratégiques ; analyse du contexte</p>	<ul style="list-style-type: none"> <li>• Cartographie des programmes des partenaires à l'aune des priorités stratégiques</li> <li>• Base de données du financement de Danida</li> <li>• Rapports des partenaires</li> <li>• Discussions à propos des défis humanitaires actuels lors d'ateliers entre parties prenantes</li> <li>• Revue documentaire sur le contexte humanitaire international</li> <li>• Entretiens avec le HCP et les partenaires</li> </ul>

Source : Mowjee *et al.* (2015: Annexe C)

Tableau 8.3 : Exemple de matrice d'évaluation pour l'évaluation d'un projet d'ONG

Critères d'évaluation	Questions-clés	Hypothèses	Sources d'information
<b>Couverture</b>	<b>Dans quelle mesure la couverture du programme de gestion communautaire de la malnutrition aiguë était-elle appropriée ?</b>	Le projet a atteint toutes les personnes de la même manière à l'intérieur des zones définies	<ul style="list-style-type: none"> <li>• Proposition de projet</li> <li>• Diagnostics</li> <li>• Enquêtes nationales</li> <li>• Rapports d'évaluation</li> <li>• Entretiens avec des informateurs clés</li> </ul>

Source : Morán (2012: Annexe 6)

Idéalement, les critères doivent apparaître seulement après les questions clés.

Tableau 8.4 : Exemples de rubriques de matrices

<b>Evaluation de l'aide alimentaire aux réfugiés de longue durée au Rwanda</b> <i>Sutter et al. (2012)</i>	<b>Evaluation de l'aide alimentaire aux réfugiés au Bangladesh</b> <i>Nielsen et al. (2015)</i>	<b>Evaluation de la réponse à la crise en RCA</b> <i>Lawday (2015)</i>
<ul style="list-style-type: none"> <li>• Questions d'évaluation</li> <li>• Sous-questions</li> <li>• Type de sous-questions</li> <li>• Mesure ou indicateurs</li> <li>• Cible ou standard (normatif)</li> <li>• Données des enquêtes de référence</li> <li>• Source des données</li> <li>• Conception</li> <li>• Echantillon ou recensement</li> <li>• Instrument de collecte des données</li> <li>• Analyse des données</li> </ul>	<ul style="list-style-type: none"> <li>• Type de questions (primaire ou secondaire)</li> <li>• Question d'évaluation</li> <li>• Questions directrices</li> <li>• Catégorie d'indicateur (préliminaire)</li> <li>• Méthodes</li> <li>• Sources d'information</li> <li>• Commentaires ou observations</li> </ul>	<ul style="list-style-type: none"> <li>• Sujet</li> <li>• Questions</li> <li>• Sous-questions</li> <li>• Critères de jugement, normes, lignes directrices, bonnes pratiques</li> <li>• Sources</li> <li>• Méthodes</li> <li>• Analyses</li> <li>• Forces/limites</li> </ul>

## 8.4 Le rapport de démarrage

Le contenu des rapports de démarrage varie en fonction du contexte et de l'échelle de l'évaluation (Tableau 8.5).

Certains sujets peuvent être inclus dans les TdR plutôt que dans le rapport de démarrage. Cela dépend de la politique suivie par le responsable de l'évaluation. Le rapport de démarrage doit permettre de clarifier les intentions de l'équipe d'évaluation aux yeux du responsable de l'évaluation. Les TdR et le rapport de démarrage doivent à eux deux décrire dans sa totalité le travail de l'évaluation.

Dans une petite évaluation, le rapport de démarrage peut couvrir seulement quelques-uns des sujets listés dans le tableau, mais il doit au minimum contenir :

- La matrice d'évaluation (où, comme c'est la norme, l'équipe doit développer ses idées)
- Le plan de travail (notamment la répartition du travail si l'équipe est de taille conséquente)
- Les outils de collecte des données qui seront utilisés

Un rapport de démarrage aussi simple conviendrait uniquement à une très petite évaluation.

Tableau 8.5 : Ce que les TdR et le rapport de démarrage doivent et peuvent contenir

<b>Sujets généralement inclus dans les TdR</b>	<ul style="list-style-type: none"> <li>• Contexte</li> <li>• Objectif général (comment l'évaluation sera utilisée)</li> <li>• Objectifs spécifiques</li> <li>• Critères</li> <li>• Périmètre</li> <li>• Public</li> <li>• Rôles et responsabilités</li> <li>• Etapes clés</li> <li>• Réalisations / produits attendus</li> <li>• Contenu du rapport de démarrage</li> </ul>
<b>Sujets qui peuvent être inclus dans les TdR ou le rapport de démarrage</b>	<ul style="list-style-type: none"> <li>• Cadre de l'évaluation</li> <li>• Questions d'évaluation</li> <li>• Sources à utiliser</li> <li>• Matrice d'évaluation</li> <li>• Conceptions de l'évaluation</li> <li>• Méthodes de collecte des données</li> <li>• Indicateurs à mesurer</li> <li>• Méthodes d'analyse des données</li> <li>• Contenu du rapport d'évaluation</li> </ul>
<b>Sujets généralement inclus dans le rapport de démarrage</b>	<ul style="list-style-type: none"> <li>• Plan de travail</li> <li>• Répartition du travail au sein de l'équipe</li> <li>• Outils de collecte des données</li> </ul>

Le rapport de démarrage doit présenter les méthodologies proposées, notamment un plan initial des entretiens prioritaires pour les entretiens à venir. Il doit faire état de tout groupe consultatif et, là où cela est possible, spécifier leur rôle. En annexe, il doit inclure un guide d'entretien et une liste des sujets abordés en *focus groups*, le cas échéant.

Il est également utile de présenter tout plan formel de diffusion dans le rapport de démarrage si l'équipe d'évaluation doit s'impliquer dans la diffusion. A titre d'exemple, si l'équipe d'évaluation doit produire un travail académique, le rapport de démarrage doit identifier quels membres de l'équipe en seront responsables.

Tableau 8.6 : Le contenu de votre rapport de démarrage



Élément	Commentaires
Contexte	<p>Cette section doit résumer le contexte dans lequel l'évaluation a lieu.</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 10px;">  <p><b>Astuce :</b> Insérer une chronologie Une chronologie peut être utile pour exposer le contexte et être complétée dans le rapport principal avec les données rassemblées durant le travail de terrain.</p> </div>
Action à évaluer	<p>Cette section doit montrer que les membres de l'équipe comprennent en quoi consiste l'action à évaluer. Il peut comprendre une description de l'intervention avec les données de base collectées par l'équipe durant la revue documentaire et inclure une théorie du changement ou tout autre modèle logique. Les tableaux de données peuvent être mis en annexe.</p>
Objectif de l'évaluation	<p>Cette section résume la compréhension que l'équipe se fait de l'objectif général et des objectifs spécifiques de l'évaluation ainsi que de l'utilisation qui en sera faite. L'équipe doit aussi y expliquer comment cela a influencé ses choix de conceptions et de méthodes.</p>
Analyse des parties prenantes	<p>Une analyse des parties prenantes peut aider l'équipe d'évaluation à planifier le travail de terrain de façon à optimiser l'utilisation des résultats de l'évaluation. Voir <a href="#">Section 3 : Penser tôt et souvent à l'utilisation de l'évaluation</a>. L'analyse des parties prenantes fournie par le rapport de démarrage de l'évaluation du Cluster Sécurité alimentaire en est un exemple (Steets <i>et al.</i>, 2013) tout comme le tableau des relations de pouvoir et d'intérêts fourni par le rapport de démarrage de l'évaluation inter-organisations en RCA (Lawday <i>et al.</i>, 2015: 14).</p>
Questions d'évaluation	<p>Cette section doit contenir toutes les versions des questions d'évaluation que l'équipe propose. Elle peut également proposer des questions retravaillées de manière à réduire le total à un nombre gérable, ou se focaliser sur les questions clés. Celles-ci sont résumées dans la matrice d'évaluation. Voir <a href="#">Section 6 : Choisir les questions d'évaluation</a>.</p>
Conception d'évaluation	<p>Alors que les conceptions sont résumées dans la matrice d'évaluation, cette section présente les raisons de votre/vos choix. Actuellement, assez peu d'EAH identifient leur conception. L'évaluation de l'aide alimentaire aux réfugiés de longue durée au Bangladesh (Nielsen, 2012: 5) fait figure d'exception. Elle désigne sa conception sous le terme <i>Post-facto non-equivalent comparison group design</i>. Voir <a href="#">Section 11 : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation</a>.</p>
Méthodes	<p>Les méthodes sont résumées dans la matrice d'évaluation, mais les méthodes que l'équipe propose d'utiliser pour collecter et analyser les données afin de répondre aux questions d'évaluation sont présentées plus en détail ici. Voir <a href="#">Section 10 : Les méthodes de recherche documentaire</a>, et <a href="#">Section 13 : Les méthodes de terrain</a>. Les détails de chaque méthode doivent indiquer la stratégie d'échantillonnage visée et l'approche proposée pour l'analyse. Les limites des méthodes de collecte des données doivent être clairement énoncées, notamment la stratégie d'échantillonnage et toute limite liée aux ressources disponibles. Voir <a href="#">Section 12 : L'échantillonnage</a>.</p>





Tableau 8.6 : Le contenu de votre rapport de démarrage

Élément	Commentaires
Outils de collecte et d'analyse des données	Ces outils sont généralement mis en annexe du rapport de démarrage et comprennent les guides d'entretien, les formulaires d'enquête, les grilles d'évaluation, ou tout autre instrument de collecte des données à utiliser. Les outils d'analyse de données doivent également être présentés, notamment toutes les listes initiales de codes. Voir <a href="#">Section 16 : L'analyse</a> .
Matrice d'évaluation	Cette matrice montre comment les évaluateurs prévoient de répondre à chaque question d'évaluation. Elle doit refléter les conceptions et méthodes exposées dans le rapport de démarrage, et peut être présentée en annexe. Alors que le texte du rapport détaille les conceptions et méthodes, la matrice montre laquelle l'équipe prévoit d'utiliser pour chaque question.
Plan de travail détaillé	Il spécifie les sites que les membres de l'équipe prévoient de visiter et quand, ainsi que les jours proposés pour les visites du siège. Il doit également indiquer les responsabilités de chaque membre d'équipe. Voir <a href="#">Section 7 : Les termes de référence et la budgétisation d'une évaluation et les éléments sur le calendrier dans la Section 9</a>
	<div style="border: 1px solid #ccc; background-color: #f9f9f9; padding: 10px;">  <p><b>Astuce :</b> Évitez les plans précis dans les environnements peu sûrs. Indiquer à l'avance des plans des itinéraires de voyage précis peut présenter un risque sécuritaire. De plus, une flexibilité de dernière minute est souvent nécessaire dans de tels environnements.</p> </div>
Plan du rapport principal, et format des autres publications	Cela prend généralement la forme d'une table des matières, et peut également inclure une description d'autres produits de l'évaluation, comme les grandes lignes d'ateliers de diffusion. Il peut également inclure un plan de diffusion détaillé. Voir <a href="#">Section 17 : Rédiger et communiquer des résultats d'évaluation en se focalisant sur l'utilisation</a> .
Cibles des entretiens	Cela fournit une liste préliminaire des personnes que l'équipe veut interroger, ou au minimum les types de personnes à interroger.
Questions et sujets en suspens	Il s'agit d'une opportunité pour mettre en avant les ambiguïtés, les sujets de préoccupation, ou les contradictions que l'équipe d'évaluation aimerait voir traités et clarifiés par l'organisation commanditaire avant la prochaine étape.
Risques et mesures de mitigation	L'équipe identifie les risques qu'elle prévoit et la manière dont elle pense les réduire.
Questions éthiques	L'équipe expose comment elle abordera toutes les questions éthiques au sein de l'évaluation. Voir <a href="#">Section 14 : Impliquer la population affectée dans votre évaluation</a> .

## Évaluer le rapport de démarrage

Le rapport de démarrage permet au responsable de l'évaluation de voir comment l'équipe comprend le travail de l'évaluation et prévoit de l'aborder. Il permet également aux parties prenantes sur le terrain de voir comment l'équipe d'évaluation prévoit de travailler de manière à ce qu'elles puissent identifier tous les sujets ou les difficultés pertinent(e)s.

Ce rapport permet également au responsable de l'évaluation de traiter ces sujets grâce à la compréhension et à l'approche qu'en a l'équipe et ce, avant qu'ils ne deviennent des problèmes majeurs. Il doit démontrer que l'équipe comprend le contexte de la crise humanitaire, le contexte de la réponse et des actions à évaluer, l'objectif et l'intention de l'évaluation ainsi que les préoccupations des parties prenantes.

Un rapport de démarrage peut être évalué à partir du [Tableau 8.6](#) des contenus possibles présenté ci-dessus et en considérant les points suivants :

- Ce rapport de démarrage doit-il contenir une telle section ?
- Cette section expose-t-elle ce qui est planifié ?
- Ce qui est exposé est-il adéquat et approprié ?

Les réponses à ces questions indiqueront les domaines dans lesquels le responsable de l'évaluation doit s'engager dans des discussions plus approfondies avec l'équipe d'évaluation.



### **Astuce :**

Cherchez à préciser ce que vous voulez. Si vous vous attendez à ce que l'équipe couvre des sujets particuliers dans le rapport de démarrage, assurez-vous de le spécifier dans les TdR.

# 9 / Planifier et gérer votre évaluation

Cette section présente différentes modalités de gestion et de gouvernance pour une évaluation. Elle aborde également les questions de calendrier de l'évaluation, de recrutement de l'équipe et de *leadership*.

L'échelle et la complexité des modalités de gestion et de gouvernance des évaluations doivent être proportionnelles à l'échelle et à la complexité de l'évaluation. Les grandes évaluations conjointes tendent à demander les modalités de gouvernance les plus complexes.

## 9.1 La décision d'évaluer

Le travail du responsable de l'évaluation commence lorsque la décision est prise de réaliser ou d'envisager de réaliser une évaluation. Voir Section 2 : Décider de réaliser une évaluation.

Cette décision peut être prise lors de la phase de conception d'un projet, durant la mise en œuvre, ou être déterminée par la politique de l'organisation en matière d'évaluation. Une politique d'évaluation peut rendre obligatoire l'évaluation de programmes d'une certaine taille ou complexité.

Dans d'autres cas, le responsable du programme, la haute direction ou le responsable de l'évaluation peuvent proposer une évaluation, soit conformément à une politique d'évaluation plus large, soit de façon ponctuelle.

## Le biais positif de l'évaluation

Les responsables de programme tendent à favoriser les évaluations de programmes qui sont considérés comme des réussites. Il peut exister une certaine résistance pour évaluer des programmes qui ont été vus comme des échecs car les responsables peuvent craindre que de telles évaluations aient un impact négatif sur leur carrière. Cela introduit un biais envers les résultats positifs dans le cadre de l'activité évaluative.

Les responsables d'évaluation, s'ils sont responsables d'un apprentissage institutionnel plus large, peuvent favoriser des évaluations de programmes vus comme relativement infructueux en raison de leur potentiel à générer des leçons. Dans de tels cas, il se peut que les responsables d'évaluation aient à demander du soutien dans le cadre de l'évaluation.



### Astuce :

Une visite de terrain réalisée par le responsable de l'évaluation avant l'évaluation peut être utile pour affiner les TdR ainsi que pour atténuer certaines inquiétudes des parties prenantes basées sur le terrain par rapport à l'évaluation (Etude de faisabilité de l'évaluation).

## 9.2 Les groupes consultatifs

Les modalités qui permettent de gérer au mieux une évaluation garantissent que les principales parties prenantes demeurent engagées dans leurs décisions. Les groupes consultatifs sont souvent formés dans ce but. Les types les plus communs sont présentés ci-dessous. Même si une évaluation de grande taille peut utiliser tous ces types de groupes consultatifs, une plus petite évaluation peut utiliser une structure de gestion beaucoup plus simple, impliquant un groupe consultatif qui comprend quelques collègues d'évaluation issus d'organisations similaires pour les fonctions de pilotage et de référence entre pairs, ainsi que le responsable d'évaluation pour les fonctions de pilotage et de gestion.



### Astuce :

Plutôt que de mettre en place un groupe consultatif, envisagez de créer un groupe de collègues d'évaluation issus d'organisations similaires afin qu'il serve de groupe consultatif permanent pour les évaluations de chaque organisation. Cela peut également constituer une opportunité pour mettre en place des partenariats avec des institutions universitaires.

## Le groupe de pilotage



### Définition : Groupe de pilotage

Groupe mis en place pour piloter une évaluation à travers ses étapes clés comme la rédaction des Termes de référence, du rapport de démarrage et du rapport final.

Dans une évaluation de grande taille, et en particulier dans le cadre d'une évaluation conjointe, un groupe de pilotage est généralement composé de responsables d'évaluation provenant de plusieurs organisations. Un *leadership* solide est essentiel pour permettre au groupe de travailler de manière efficace et d'exploiter pleinement son potentiel, en particulier si les membres proviennent d'horizons divers. Dans une évaluation de plus petite taille, le comité de pilotage de l'évaluation peut comprendre le responsable de l'évaluation et un ou deux collègues. Le groupe de pilotage se réunit uniquement au moment des étapes clés de l'évaluation.

L'évaluation de la réponse à l'urgence de 2002-2003 en Ethiopie a été réalisée par un comité de pilotage comprenant des représentants du gouvernement, des bailleurs, des Nations unies et des ONG (Simpkin *et al.*, 2004: 4). Pour l'évaluation de la politique de transversalisation âge, genre et diversité du HCR, le comité de pilotage du personnel du HCR, des gouvernements et des ONG « s'est réuni trois fois à Genève. Il a revu et validé les méthodes d'évaluation, revu les rapports intermédiaires, et a commenté les conclusions et recommandations » (Thomas et Beck, 2010: 10).

Les comités de pilotage sont également utilisés pour des évaluations de petite taille. L'évaluation Care–Save the Children de la réponse au tremblement de terre en Haïti disposait d'un comité de pilotage comprenant un représentant de chaque organisation (O'Hagan *et al.*, 2011 37). Dans ce cas, le comité de pilotage a choisi les sites des visites de terrain (*ibid.* : 13). Même si on les retrouve le plus souvent dans les évaluations conjointes, les comités de pilotage peuvent également être utilisés pour des évaluations d'une seule organisation. L'évaluation de l'information, de l'assistance juridique et du conseil du NRC en Afghanistan et au Pakistan avait un comité de pilotage composé de cinq membres : le Directeur régional et quatre employés du siège (Pierce, 2009: 10).



## Le groupe de gestion

### Définition : Le groupe de gestion

Groupe qui gère l'évaluation au jour le jour et dont les tâches comprennent la rédaction des Termes de référence, le recrutement et la gestion de l'équipe d'évaluation, la gestion de la relecture et la finalisation du rapport d'évaluation.

Dans le cas d'un groupe de pilotage de grande taille, il est conseillé de mettre en place un plus petit groupe de gestion qui peut rapidement prendre des décisions si nécessaire sans avoir besoin de convoquer une réunion de l'ensemble du groupe de pilotage. De manière générale, le groupe de gestion se réunit plus souvent que le groupe de pilotage au complet. Les évaluations de plus petite taille peuvent avoir recours à un groupe de gestion plutôt qu'à un comité de pilotage. Une petite évaluation peut être réalisée par le responsable de l'évaluation sans qu'il soit nécessaire de constituer un comité de pilotage ou un groupe de gestion.

L'ETR de trois mois portant sur la réponse au séisme de 2010 en Haïti disposait d'un groupe de gestion comprenant des représentants d'OCHA, d'UNICEF et d'une ONG (Grünewald *et al.*, 2010: 19). Dans l'ETR de la réponse de l'UNICEF à la crise malienne, les TdR exigeaient que le groupe de gestion se réunisse chaque semaine pour examiner les progrès (Leonardi et Arqués, 2013: 90). La structure du rapport d'évaluation a été établie d'un commun accord entre l'équipe et le groupe de gestion (*ibid*: 24).

Dans le cadre des évaluations de petite taille, le responsable de l'évaluation peut accomplir toutes ces tâches.

## Le groupe de référence



### Définition : Le groupe de référence

Groupe constitué des principales parties prenantes connaissant bien l'environnement local et pouvant apporter des conseils sur des questions pratiques en lien avec l'évaluation et sur la faisabilité des recommandations qui en découlent.

Alors que les termes « comité de pilotage » et « groupe de gestion » sont utilisés de façon assez constante, c'est moins vrai pour les termes « groupe de référence » et « groupe de revue par les pairs ». Le groupe de référence d'une organisation correspond parfois au groupe de revue par les pairs d'une autre organisation. Les normes d'évaluation d'UNEG renvoient à un « examen par des pairs, ou par un groupe de référence, composé d'experts extérieurs » (UNEG, 2005: 15). Le PAM distingue le groupe de référence interne des membres de la revue par des pairs externes. L'évaluation stratégique de la planification d'urgence du PAM disposait d'un groupe de référence interne composé de six membres et d'un groupe de revue par des pairs externes comprenant trois membres (Ressler *et al.*, 2009: ii). Le rapport de

l'évaluation de l'engagement danois en Somalie et dans les pays avoisinants (Gardner *et al.*, 2012: 18) a été révisé suite à des commentaires formulés par le groupe de référence.

Établir un groupe de référence est une bonne façon d'impliquer les principales parties prenantes lorsqu'une évaluation a été commanditée par le siège. L'évaluation du programme éducatif de l'UNICEF au Timor oriental disposait d'un groupe de référence composé de parties prenantes clés identifiées par le bureau pays (Tolani-Brown *et al.*, 2010: 93)



### **Exemple de bonnes pratiques : Optimiser l'implication du bureau pays dans l'évaluation**

Pour rendre l'exercice aussi utile que possible au niveau pays, le Bureau de l'évaluation de l'UNICEF a cherché à impliquer le bureau pays dans toutes les phases de l'évaluation en créant un groupe de référence local. Le responsable de l'évaluation s'est rendu au Timor oriental et a impliqué le responsable suivi-évaluation (S&E) du bureau pays dans des discussions portant sur la stratégie d'échantillonnage et la conception de l'enquête, tout en offrant un appui au niveau de l'assurance qualité de l'évaluation afin d'améliorer la qualité générale de l'évaluation. Cet élément de renforcement des capacités a servi d'incitation pour que le personnel S&E soutienne de façon active le programme éducatif et a permis que ces employés aient un rôle dans son succès. L'évaluation s'est avérée de très bonne qualité et pertinente, donnant lieu à une forte réponse et à un important processus de suivi mené par le bureau pays. Grâce à cette approche indépendante mais collaborative, le bailleur a engagé une seconde tranche importante du financement pluriannuel en faveur du programme. Ce processus a également modifié la culture du bureau pays qui est devenue beaucoup plus favorable aux évaluations.

Source : UNICEF (2010)

## **Le groupe de revue par des pairs**



### **Définition : Le groupe de revue par des pairs**

Groupe qui conseille sur les questions de qualité, généralement composé d'évaluateurs et d'autres spécialistes choisis pour leurs connaissances de l'évaluation, de la région ou du type d'intervention évaluée.

Il n'est pas nécessaire qu'un groupe de revue par les pairs comporte beaucoup de personnes. Il peut apporter des conseils sur l'analyse contextuelle, la conception et la méthodologie, mais aussi commenter la rédaction du rapport. Il peut s'avérer particulièrement utile pour des évaluations gérées par un directeur général plutôt que par un évaluateur, et lorsque l'équipe d'évaluation ne connaît pas bien la région. L'étude de la réponse à la crise du Kosovo de 1999 réalisée par le Disasters Emergency Committee (DEC) a eu recours à une équipe de revue par les pairs composée d'experts régionaux et humanitaires qui a donné des instructions à l'équipe d'évaluation avant qu'elle ne parte sur le terrain, puis a commenté le rapport avec eux (Wiles *et al.*, 2000: 4). L'évaluation de la coopération de la FAO en Somalie a remarqué que les deux membres de la revue par les pairs avaient fourni « des apports et des commentaires constructifs et judicieux » (Buchanan-Smith *et al.*, 2013: v). L'évaluation a également été sujette à un processus de revue par les pairs au sein du département d'évaluation (*ibid.*: 4).

Le nombre de membres d'une revue par des pairs varie. L'évaluation des programmes de sécurité alimentaire et de moyens de subsistance en ville d'Oxfam GB à Nairobi (Macaulan et Phelps, 2012) comprenait un seul examinateur pair. L'évaluation du DEC portant sur les inondations au Mozambique en 2000 en avait trois (Cosgrave *et al.*, 2001 p.1) : le chef d'équipe leur soumettait la version provisoire du rapport d'évaluation ainsi qu'aux autres membres de l'équipe (*ibid.* p.107). Les pairs examinateurs commentaient également la version finale.



**Astuce :**

Séparez le groupe de revue par des pairs et le groupe de référence afin que chacun puisse se concentrer sur la tâche qui lui a été assignée sans être distrait. Même si leurs responsabilités semblent se recouper, ils fonctionnent mieux de manière séparée.



**Astuce :**

Incluez dans le budget de l'évaluation un perdiem pour les pairs examinateurs qui travaillent sur leur temps libre et non pas en tant que salariés.

Les évaluations conjointes de taille importante tireront des bénéfices de la présence d'une grande diversité de groupes consultatifs, mais des mécanismes plus simples conviendront mieux à des évaluations de plus petite taille. Même la plus petite évaluation peut tirer profit d'un groupe de référence des parties prenantes et même d'un seul pair examinateur.



## 9.3 Le calendrier de l'évaluation

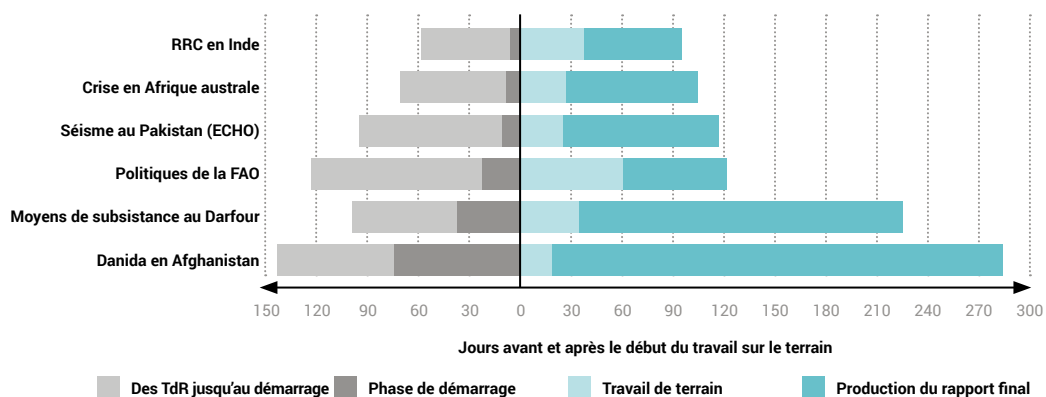
La longueur d'une évaluation dépend du contexte.  
Le tableau suivant donne certaines estimations.

Tableau 9.1 : Calendrier indicatif d'une évaluation

Activité	
<b>Développer les <u>TdR</u></b>	D'une semaine à plus d'un an selon le niveau de consultation et de validation nécessaire. Le IASC a développé un modèle de TdR qui peut être facilement adapté (IASC, 2010). Cette approche convient particulièrement lorsqu'une série d'évaluations similaires est planifiée, ou dans le cas d'une organisation se concentrant sur un point en particulier (par exemple, une focalisation sur les enfants ou une comparaison d'organisations).
<b>De la publication des <u>TdR</u> à la décision de l'attribution</b>	Cela peut aller assez vite dans le cas d'une évaluation interne si le personnel est déjà disponible. Cela prend plus de temps lorsque des consultants externes sont engagés. Graig (2013) suggère que trois à quatre semaines suffisent pour que les évaluateurs soumettent une manifestation d'intérêt, mais cela peut prendre plus de temps s'il s'agit d'un large appel à manifestation d'intérêt. Il est plus courant de voir ce processus et la désignation du vainqueur prendre deux mois.
<b>Formalités contractuelles</b>	D'un jour à plusieurs mois, en fonction de l'organisation.
<b><u>Phase de démarrage</u></b>	D'une semaine à plusieurs mois, voire plus. Des phases de démarrage plus longues sont nécessaires dès lors qu'une visite de démarrage et/ou une revue documentaire approfondie est nécessaire.
<b><u>Travail de terrain</u></b>	D'une semaine (pour les plus petites évaluations) jusqu'à plusieurs mois, en fonction de l'échelle et de la complexité. La durée la plus courante pour les EAH est de deux à quatre semaines.
<b><u>Analyse des données et rédaction du rapport</u></b>	D'une semaine à plusieurs mois, en fonction de la complexité. Une analyse des données de l'enquête peut prendre plusieurs mois s'il est nécessaire de traduire des données collectées ou de coder des données.
<b>Cycle de révision</b>	Chaque cycle de révision comprend une phase de commentaires, généralement de deux semaines pour une petite évaluation et jusqu'à quatre semaines ou plus pour une évaluation plus complexe où les organisations doivent consulter leur personnel sur le terrain. Cette phase est suivie par une à deux semaines de révision du rapport provisoire. Presque tous les plans d'évaluation nécessitent un seul cycle de révision, mais il peut y en avoir beaucoup plus, en particulier lorsque les résultats de l'évaluation sont moins bien accueillis ou qu'il est nécessaire de parvenir à un consensus.
<b><u>Activité de diffusion</u></b>	Elle peut comprendre des ateliers ou des visites de terrain pour diffuser les résultats, généralement dans le mois qui suit la publication du rapport.

Le diagramme ci-dessous donne la chronologie de huit évaluations de différents types dans lesquelles les auteurs ont été impliqués. La plupart étaient des évaluations de taille intermédiaire. Les plus petites évaluations étaient rapides que les grandes.

**Schéma 9.1** : Calendriers d'évaluations mentionnant en jours les périodes situées avant et après le début du travail sur le terrain



Des raisons particulières expliquent pourquoi les différentes phases ont pris autant de temps dans chaque cas. Cette illustration sert ainsi à montrer l'ampleur des variations et non à présenter des normes.

## 9.4 Les évaluations internes et externes

Une évaluation interne est réalisée par le personnel de l'organisation qui est évaluée. Certaines organisations utilisent les termes interne et externe pour renvoyer non pas à l'organisation mais à son programme, par exemple lorsqu'une évaluation conduite par du personnel de l'organisation non impliqué dans le programme est désignée comme une évaluation externe. Une évaluation externe est réalisée par une équipe d'évaluation qui est extérieure à l'organisation qui est évaluée.

Certaines organisations ont des départements Evaluation dont le personnel a des parcours professionnels atypiques et rapporte directement à la direction plutôt qu'aux responsables de programme. Ces organisations avancent que, dans de telles circonstances, même une évaluation interne peut être indépendante.

Dans certains cas, une évaluation peut être conduite par une équipe combinant consultants externes et personnel interne. C'est une évaluation mixte. Si la personne qui dirige l'évaluation est externe, l'évaluation est considérée comme indépendante. La combinaison d'un chef d'équipe externe et de personnel interne peut faciliter l'apprentissage.

Le choix d'une équipe d'évaluation externe, interne ou mixte dépend de l'objectif de l'évaluation (Voir [Section 2 : Décider de réaliser une évaluation](#)). Si l'évaluation vise principalement la redevabilité, les évaluateurs doivent être externes parce des évaluateurs internes peuvent ne pas être vus comme suffisamment indépendants. En fonction du contexte organisationnel, il peut être préférable d'inclure un ou deux employés dans l'équipe – provenant éventuellement du département Evaluation de l'organisation, s'il existe, ou, si cela est pertinent, des départements opérationnels. Ces membres ne doivent jamais être majoritaires ou assumer le *leadership*, l'arrivée de tout personnel interne dans l'équipe pouvant mettre en danger la perception liée à l'indépendance de l'évaluation.

Si l'objectif principal de l'évaluation est l'apprentissage, l'équipe d'évaluation doit comprendre une majorité de personnel interne (qui est supposé bénéficier des processus d'apprentissage) ou engager une équipe d'évaluateurs externes dont le rôle principal sera de faciliter l'apprentissage de l'équipe. Les équipes mixtes peuvent aider à boucler la boucle de l'apprentissage car le personnel de l'organisation apportera des éclairages sur la meilleure manière de mettre en œuvre l'apprentissage organisationnel. A titre d'exemple, ce personnel peut être au courant de parties prenantes clés à impliquer et de domaines organisationnels ouverts au changement (Riccardo Polastro, communication personnelle, juin 2013).



**Question clé : Indépendant ?**

Les prestataires extérieurs sont vus comme plus indépendants que le personnel de l'organisation par rapport aux pressions de la direction pour ce qui est de présenter un programme sous un jour favorable. Cela est vrai de manière générale, mais là où un évaluateur ou une entreprise d'évaluation espère continuer à travailler avec un client, quel impact cela peut-il avoir sur leur indépendance ?

Les avantages et inconvénients propres à l'utilisation d'évaluateurs internes et externes sont présentés dans le [Tableau 9.2](#).

Tableau 9.2 : Avantages et inconvénients des évaluateurs internes et externes

Évaluateurs internes		Évaluateurs externes	
+	Les bénéfices tirés de l'apprentissage et des réflexions nés durant le processus d'évaluation restent au sein de l'organisation.	+	Ils sont souvent plus objectifs.
+	Ils connaissent l'organisation et sa culture.	+	Ils risquent moins d'être influencés par l'organisation.
+	Ils sont connus du personnel.	+	Ils apportent un regard neuf.
+	Ils peuvent être moins menaçants et bénéficier d'une plus grande confiance.	+	Ils peuvent avoir une plus grande expérience sur laquelle s'appuyer.
+	Les résultats et recommandations peuvent être plus adaptés à l'organisation.	+	Ils peuvent consacrer plus de temps à l'évaluation.
+	Les recommandations ont souvent beaucoup plus de chance d'être adoptées.	+	Ils peuvent servir d'experts ou de facilitateurs extérieurs.
+	Ils coûtent moins cher.	+	Ils ne font pas partie de la structure de pouvoir de l'organisation.
+	Cette option renforce la capacité d'évaluation interne et contribue, de manière générale, au renforcement de la capacité du programme.	+	Ils peuvent apporter des ressources supplémentaires.
-	Leur objectivité peut être remise en question.	+	Ils sont plus susceptibles d'être formés et expérimentés en matière d'évaluation.
-	La structure organisationnelle peut limiter leur participation.	+	Ils sont considérés comme des experts.
-	Leurs engagements professionnels peuvent limiter leur participation.	-	Ils peuvent ne pas connaître l'organisation.
-	Leur motivation peut être questionnée.	-	Ils peuvent ne pas connaître les contraintes qui affecteront l'adoption des recommandations.
-	Ils peuvent accepter trop facilement les postulats de l'organisation.	-	Les bénéfices tirés de l'apprentissage et des réflexions nés durant le processus d'évaluation ne restent pas au sein de l'organisation.
-	Ils ne sont peut-être pas formés aux méthodes d'évaluation.	-	Ils peuvent être perçus comme des adversaires.
-	Ils peuvent réduire la crédibilité de l'évaluation à l'extérieur de l'organisation.	-	Ils coûtent plus cher.
-	Ils peuvent avoir des difficultés à éviter certains biais	-	Les embaucher peut demander de longues négociations contractuelles.
-	Ils peuvent manquer d'expertise technique.	-	Ils peuvent rendre le suivi des recommandations plus incertain.
		-	Ils peuvent ne pas connaître l'environnement.
		-	Ils peuvent être influencés par leur besoin de s'assurer de nouveaux contrats et, par conséquent, être moins indépendants qu'ils n'en ont l'air.

Compilation des auteurs. Voir Gosling et Edwards (2003) ainsi que CRS et Short Cut on Preparing for an Evaluation (2008) de la Croix-Rouge américaine.

## L'apprentissage des parties prenantes dans le cadre des évaluations

La réalisation d'une évaluation fournit à l'équipe d'évaluation de grandes opportunités en termes d'apprentissage. Si l'apprentissage des parties prenantes est la priorité, une évaluation externe peut ne pas être la meilleure approche. Les évaluations internes permettent au personnel de l'organisation de saisir pleinement l'apprentissage, augmentant ainsi la probabilité que les leçons tirées soient utilisées. Des évaluations mixtes, éventuellement menées par un évaluateur externe expérimenté avec quelques employés de l'organisation, constituent une bonne opportunité pour l'apprentissage tout en préservant la qualité et la crédibilité de l'évaluation.

Une autre approche est celle du modèle de l'évaluation par les pairs. Elle peut être réalisée soit avec :

- Des organisations partenaires qui participent à l'évaluation d'un partenaire.
- Des programmes de différents pays qui participent à l'évaluation d'un programme similaire dans un autre pays.

Cette approche garantit que les personnes qui tirent les leçons de l'évaluation occupent des postes où elles peuvent appliquer cet apprentissage dans d'autres contextes. Encore une fois, employer un évaluateur externe expérimenté aide à assurer la qualité et la crédibilité de l'évaluation.

## 9.5 Recruter les évaluateurs

Les évaluateurs externes doivent faire l'objet de contrats. Pour cela, les responsables de l'évaluation ont plusieurs options :

- Publier un appel d'offres formel comprenant les détails de l'évaluation sous la forme de TdR joints ou comme une partie de l'appel d'offres.
- Publier des TdR comprenant les détails du contrat.
- Publier une lettre d'intention exposant les raisons qui amènent l'organisation à commanditer une évaluation.

Certaines organisations utilisent le terme « lettre d'intention » pour signifier leur expression d'intérêt. Dans ce Guide, nous l'utilisons pour faire référence à la proposition que remettent les évaluateurs à une organisation pour manifester leur intérêt à réaliser une évaluation.

Davidson (2012) et Graig (2011 ; 2013) préconisent le recours à un processus d'expression d'intérêt pour sélectionner l'évaluateur ou identifier des contractants potentiels. Recourir à une expression d'intérêt pour sélectionner des évaluateurs est particulièrement approprié pour les évaluations de plus petite taille. Davidson (2010b) suggère que l'on pose les questions suivantes aux personnes remettant une expression d'intérêt :

- Quelle est l'entité/la personne/l'équipe qui envisage de soumettre une offre pour l'évaluation? (1 page maximum)
- Pourquoi cette évaluation l'intéresse-t-elle ? (une demi-page maximum)
- Que pensent-ils pouvoir faire ? De quelle expertise, quelle expérience et quelles capacités adéquates disposent-ils ? (1 page maximum)
- Qui sont-ils – une entreprise spécialisée dans l'évaluation, un évaluateur, ou une équipe d'évaluation ? Quels sont leurs valeurs distinctives, leurs pratiques, domaines d'expertise et spécialisations ? (une demi-page maximum)
- Quels sont les tarifs journaliers des membres de l'équipe proposée ? Quels frais de gestion ou imprévus demanderaient-ils ?
- Si possible, fournir deux ou trois résumés exécutifs de rapports d'évaluation récents que le/les évaluateur(s) chef d'équipe ont dirigés.

Davidson suggère en outre de ne pas demander le CV des évaluateurs ou d'autres documents durant cette phase de présélection. Les CV peuvent être appropriés lorsqu'un contrat sera uniquement basé sur des expressions d'intérêt, comme cela peut être le cas pour des évaluations de petite taille.

Davidson remarque que les demandes standard de proposition donnent lieu à des offres pleines de détails sans intérêt, sans grande différence de l'une à l'autre. Elle suggère que les évaluateurs soient sélectionnés non à partir de leur plan d'évaluation mais de leur capacité à traiter les défis majeurs. Parmi les points à questionner, on peut citer :

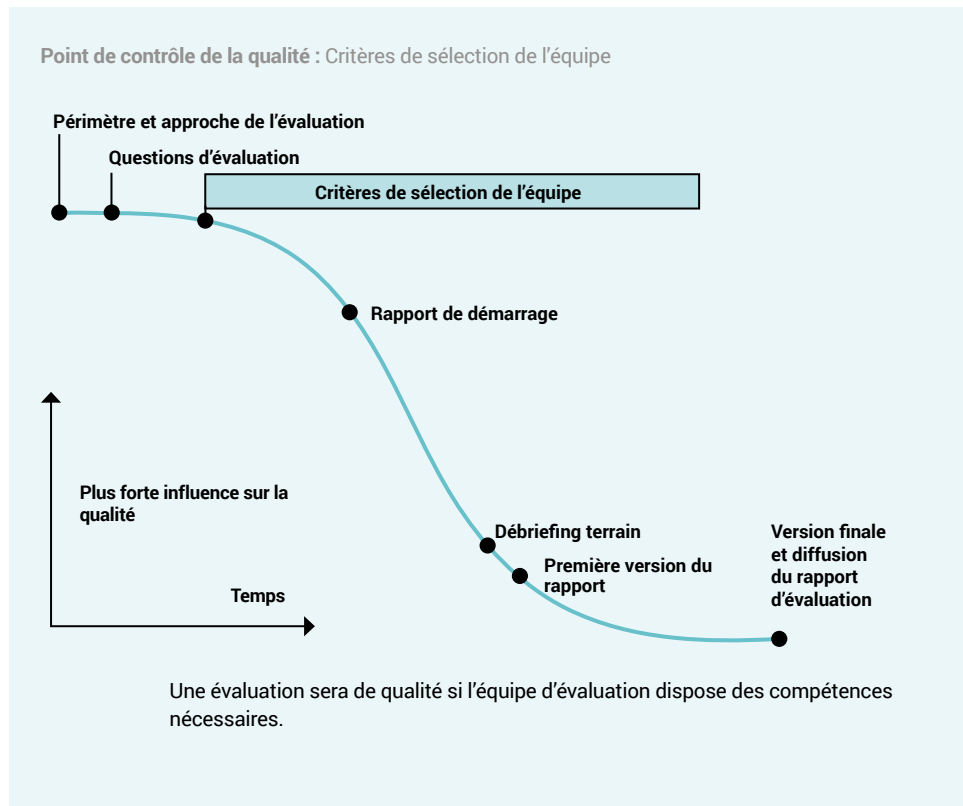
- S'il s'agit d'une évaluation plutôt qu'un travail de recherche sur ce sujet, demandez-leur d'expliquer la différence. Quelles sont les connaissances et les compétences qu'ils pensent devoir appliquer à une évaluation plutôt qu'à un projet de recherche ?
- Comment répondraient-ils à une question du type « Dans quelle mesure les résultats sont-ils utiles pour [un groupe spécifique de bénéficiaire?] »
- Comment gèreraient-ils des points de vue divergents sur les résultats les plus utiles et sauraient-ils à quoi correspond la qualité ? Comment les appliqueraient-ils de manière pertinente pour tirer des conclusions ?
- Comment gèreraient-ils tout défi anticipé de l'évaluation ? Demandez comment ils ont fait par le passé dans des cas similaires.

- Comment se décriraient-ils en tant qu'évaluateurs professionnels ? Quelle est leur « marque de fabrique » ? Quelle est leur approche personnelle, celle pour laquelle ils sont les mieux connus ? Quels types de projet met le mieux en valeur leurs forces ? Quels types de projet évitent-ils parce qu'ils sont en dehors de leurs points forts ? Faites des recherches sur leur expertise et leur approche en matière d'évaluation ; ne vous contentez pas de leur réponse sur leur expertise au niveau du contenu.
- Demandez-leur trois résumés exécutifs d'évaluations récemment terminées. Cela en dit beaucoup sur leur capacité à rédiger, à aller à l'essentiel, à être succinct, à répondre aux questions évaluatives importantes (et non pas seulement à les associer à des données), mais aussi sur leur capacité à vraiment comprendre les utilisateurs visés et ce qui a du sens pour eux (Davidson, 2012b).

Davidson (2010c) suggère que les réponses à ces questions prennent la forme d'un document écrit ou d'une présentation orale. Ces questions ou des questions similaires peuvent servir de base à un entretien (face-à-face ou à distance) avec le chef d'équipe. Selon Davidson, cet entretien permettrait de d'aborder une bonne partie des points techniques.

## Composer la meilleure équipe externe possible

La définition des critères de sélection de la proposition gagnante aura une forte influence sur la qualité de l'évaluation. Ces critères peuvent être présentés dans les TdR ou dans un appel d'offres distinct.



Les critères de sélection doivent garantir que l'équipe a :

- La connaissance nécessaire du contexte
- Les compétences nécessaires en matière de communication
- Les compétences pour réaliser un travail sur le terrain de haute qualité
- Les compétences analytiques permettant de convertir les éléments de preuve issus du terrain en des réponses convaincantes aux questions d'évaluation.



Critères qualitatifs de sélection	Nb max. de points
Comprendre les TdR et l'objectif des services à fournir	10
Méthodes pour la structuration de l'évaluation, méthodes de collecte et d'analyse des données	40
Organisation des tâches et de l'équipe, calendrier	10
Adéquation de l'équipe par rapport à l'expertise proposée	40
<b>Total des points techniques</b>	<b>100</b>
<b>Facteur financier : prix de la proposition la moins chère /prix de cette proposition</b>	
<b>Total général : total des points techniques multiplié par le facteur financier</b>	

(D'après Maunder *et al.* (2015))

Le tableau suivant est un exemple des critères de sélection des propositions couramment utilisés pour les évaluations de la Commission européenne. Dans cet exemple, 40 % des points techniques renvoient à la composition de l'équipe.

Sans une décomposition plus précise du sens de chaque critère, la sélection demeure quelque peu subjective. L'exemple qui suit (librement inspiré de l'évaluation des programmes de Swiss Solidarity en Haïti, voir Groupe URD, 2013) est une approche plus structurée.

Aspect	Critère	Points	
<b>Chef d'équipe</b>	Expérience significative dans le domaine de l'évaluation humanitaire, en utilisant des méthodes variées	15	
	Maîtrise du français ou du créole haïtien	5	
	Excellentes compétences rédactionnelles en français et en anglais	10	
	Une grande expérience en Haïti est un atout	5	
<b>L'équipe dans son ensemble</b>	Expérience en matière d'évaluations d'interventions de réhabilitation et de développement	5	
	Forte expertise sur Haïti, notamment une compréhension approfondie des structures économiques, des montages institutionnels et du tissu social du pays	10	
	Bonne connaissance de la langue française	10	
	Capacité à analyser des données quantitatives dans le cadre d'enquêtes (ce critère peut être rempli par un membre de l'équipe qui ne se rend pas sur le terrain)	5	
	Expérience en matière d'interventions de renforcement des compétences ou dans l'évaluation de ce type de programmes	5	
	Expérience en matière de programmes axés sur les moyens de subsistance ou dans l'évaluation d'interventions axées sur les moyens de subsistance	5	
	Expérience en matière de programmes RRC	5	
	Une bonne connaissance du créole haïtien est un atout	5	
	<b>Approche</b>	Dans quelle mesure l'approche proposée démontre une bonne compréhension des risques et des contraintes	10
	<b>Soumissionnaire</b>	Assurance d'une composition d'équipe stable sur toute la durée de l'évaluation	10
Expérience dans des évaluations similaires		10	
<b>Prix</b>	Score = 15x (coût de la proposition la moins chère)/ (coût de cette proposition)	15	
<b>Score maximum possible</b>		130	

Cette série de critères est beaucoup plus détaillée, ce qui facilite considérablement la notation. La plupart des points techniques (85/115, soit 74 %) concernent les compétences de l'équipe. Même si vous ne souhaitez pas ce niveau de détail, il est utile de spécifier les compétences demandées au chef d'équipe et à l'équipe. Cet exemple peut encore être amélioré en donnant des exemples de scores élevés et de scores bas comme :

- Capacité à saluer en créole, 1 point, capacité à mener des entretiens en créole sans aide et sans incompréhensions, 5 points.
- Expérience d'Haïti, 1 point par année d'expérience.
- Maîtrise de la langue française : allant de la capacité à « se débrouiller », 1 point ; jusqu'à la langue maternelle, 10 points.

Lorsqu'une équipe comprend un grand nombre de personnes, elles peuvent être notées individuellement avant de calculer la moyenne, à moins qu'il suffise qu'un seul membre dispose de la compétence requise.

Aspect	Critère	Nb total de points	Minimum pour se qualifier
<b>Chef d'équipe</b>	Expérience et compétences du chef d'équipe, décomposées en quelques éléments clés comme l'expérience de direction d'évaluations, l'expérience dans la zone géographique, l'expérience par rapport au secteur ou à l'organisation, etc., pour une plus grande transparence, il peut être pertinent de prévoir une sous-note pour les éléments individuels, (par exemple, l'expérience en tant que chef d'évaluations humanitaires – 1 point pour chaque tranche de deux ans jusqu'à un maximum de 7 points - ou les compétences linguistiques avec 4 pour langue maternelle, 3 pour courant, 2 pour semi-courant et 1 pour niveau de base).	20	15
<b>Entretien avec le chef d'équipe</b>	Entretien pour définir la direction de l'équipe et les compétences interpersonnelles.	20	15
<b>Autres membres</b>	Expérience et compétences des autres membres de l'équipe, décomposées en domaines d'expérience et compétences clés.	30	20
<b>Contexte</b>	Compréhension démontrée du contexte et des problèmes clés liés à l'évaluation.	10	5
<b>Approche</b>	Approche proposée pour le travail de l'évaluation (sinon, cela peut faire partie de l'entretien avec le chef d'équipe).	20	10
<b>Score technique total / score minimum pour se qualifier</b>		100	65

**Astuce :**

Tous les critères n'ont pas la même importance. Déterminez parmi vos critères de sélection ceux qui sont essentiels pour garantir une bonne évaluation.

Il est possible de définir des critères essentiels en appliquant une note minimum pour chaque critère pertinent. Par exemple, on peut décider de prendre en compte seulement les chefs d'équipe dont le score est supérieur à 10 sur un maximum de 15 pour les précédentes expériences d'évaluation. Une note élevée à un éventuel entretien de sélection (reflétant les compétences interpersonnelles) ou des références favorables de précédents clients peuvent être autant de critères essentiels.

Les points relatifs au contexte et à l'approche peuvent être scindés entre l'expression d'intérêt et l'entretien avec le chef d'équipe lorsque la décision est prise sur ces éléments plutôt que sur une proposition complète.

Voir la sous-section suivante sur les procédures contractuelles pour une discussion sur l'impact de la question du coût et la raison pour laquelle la proposition la moins chère ne doit pas automatiquement être retenue.<sup>1</sup>

**Astuce :**

Cela prend du temps de trouver le(s) bon(s) consultant(s) pour une évaluation mais cela revêt une extrême importance. Plus vous recrutez en avance, plus vous aurez le choix.

Le type de consultant varie en fonction du travail nécessité par l'évaluation. Pour pouvoir gérer des questions sensibles de partenariats, les consultants devront comprendre les relations institutionnelles en général et les partenariats en particulier. Une ETR requiert des consultants dotés d'une expérience opérationnelle suffisante pour être crédible avec le personnel de terrain. Une évaluation sectorielle peut nécessiter des évaluateurs ayant une expérience de travail dans le secteur en question. Ces exigences devraient être reflétées dans les critères de sélection de l'équipe d'évaluation.

Parmi les éléments à prendre en compte pour choisir les consultants, on peut citer :

- **Dans quelle mesure le consultant connaît vraiment l'organisation.** Un consultant qui connaît déjà l'organisation passera moins de temps à se renseigner sur elle mais pourrait être moins à même d'en remettre en question les croyances dominantes.
- **Dans quelle mesure le consultant connaît le contexte.** Ceci est particulièrement important pour les crises et les conflits complexes. Selon l'évaluation, il peut être plus important de comprendre le contexte ou le type d'activité à évaluer.

- **Savoir si le consultant a les compétences requises pour les conceptions et méthodes d'évaluation proposées et l'expérience en la matière.**  
Cela peut comprendre la connaissance des domaines sectoriels (comme l'eau-assainissement-hygiène) et la théorie appropriée (comme la théorie des organisations et l'ingénierie des processus pour les évaluations de processus organisationnels).
- **La composition globale de l'équipe.** Les équipes de plus grande taille peuvent offrir plus de compétences mais peuvent également coûter plus cher, demandent un soutien logistique plus important sur le terrain, comportent de plus grands risques de désaccords internes, et prennent plus de temps pour rédiger un rapport.

Une autre question est celle du recrutement des consultants de façon individuelle ou comme une équipe via une société de consultants. Les avantages et les inconvénients de chaque choix sont résumés dans le [Tableau 9.3](#). Dans de nombreux cas, le choix dépendra des politiques de l'organisation, ou des conséquences sur le calendrier.<sup>2</sup>

**Tableau 9.3 :** Avantages et inconvénients à engager des consultants de façon individuelle ou en équipe

Individus		Equipe de consultants	
+	Vous sélectionnez exactement les personnes et les compétences que vous souhaitez.	+	Il n'y a qu'un seul contrat à gérer.
+	Vous pouvez vous assurer que tous les membres de l'équipe ont quelque chose à offrir.	+	Vous avez plus d'influence sur les questions de qualité.
+	Peut être moins complexe qu'un contrat.	+	La résolution des conflits au sein de l'équipe est de la responsabilité des consultants.
+	Vous bénéficiez d'une relation directe avec chaque membre de l'équipe.	+	Vous ne devez gérer qu'un petit nombre de paiements.
-	Cette option est en général moins coûteuse.	-	Le contractant assume une part des risques.
-	Des limitations financières peuvent vous empêcher d'embaucher les consultants les plus qualifiés.	-	Cette option est généralement plus coûteuse.
-	Il est de votre responsabilité de résoudre les conflits au sein de l'équipe.	-	L'équipe peut parfois être composée d'un seul expert confirmé entouré d'une équipe de consultants « junior ».
-	Il vous faut fournir des ressources humaines en soutien à l'équipe.	-	Certains membres de l'équipe peuvent ne jamais avoir travaillé ensemble auparavant.
-	Vous devez organiser la logistique sur le terrain.	-	Il vous faudra parfois gérer la logistique sur le terrain et d'autres tâches si le contractant n'est pas compétent.
-	Vous devez gérer plusieurs paiements différents.	-	Des limitations financières peuvent vous empêcher d'embaucher une équipe dans le temps opportun.
-	Vous assumez tous les risques encourus.		

Source : Compilation des auteurs.

Il est parfois possible de sélectionner des consultants sur la base de précédentes expériences en commun, mais ce luxe n'est pas toujours permis. Vous pouvez demander des précisions sur l'approche qu'ils souhaitent adopter, mais il est préférable de sélectionner les consultants sur la base de leurs précédents travaux. Qui a écrit les rapports d'évaluation que vous considérez de bonne qualité ?



**Astuce :**

Ne vous fiez pas uniquement au potentiel statut d'auteur d'un consultant lors d'une précédente évaluation, contactez le responsable d'évaluation et renseignez-vous sur les antécédents professionnels du consultant. Il est possible que le rapport final soit en réalité la 19<sup>e</sup> version et ait été complètement réécrit par un autre membre de l'équipe.

La relation contractuelle dépendra de la politique de recrutement de l'organisation. Une pratique courante consiste à lier les paiements aux différents phases de l'évaluation, comme la signature du contrat, la validation du rapport de démarrage, la fin du travail sur le terrain et l'acceptation des diverses versions et du rapport final. Davidson (2012a) note que les termes du contrat doivent permettre au responsable de l'évaluation de mettre un terme au contrat si l'équipe d'évaluation s'avère incapable de respecter ses obligations.



**Astuce :**

Mentionnez explicitement dans les termes du contrat le droit de clore l'évaluation à la phase de démarrage si le rapport de démarrage n'est pas satisfaisant.

## Les procédures contractuelles

Si les évaluateurs doivent soumettre une offre, la pratique habituelle consiste à demander que l'offre technique et l'offre financière soient remises dans des enveloppes scellées distinctes. Les offres techniques sont ouvertes en premier, et les points attribués à chaque offre (voir ci-dessous la sous-section [Composer la meilleure équipe externe pour un exemple des critères qui peuvent être spécifiés](#)). Les évaluateurs peuvent ensuite être notés pour l'entretien. Une fois que les points techniques ont été attribués, les offres financières sont ouvertes et la note totale est calculée.

Si l'évaluation est attribuée sur la base d'une expression d'intérêt et d'un entretien avec le chef d'équipe, il faut demander aux candidats de remettre une offre financière scellée qui ne sera ouverte qu'après l'entretien. Cette approche est particulièrement adéquate pour les évaluations de plus petite taille.

Il existe trois grandes façons d'attribuer les contrats :

- À l'offre la moins chère parmi toutes celles qui ont été retenues sur le plan technique. Les offres qui obtiennent un niveau minimum défini de points techniques et qui ne sont disqualifiées par aucun critère clé (par exemple, obtenir zéro point pour un critère essentiel comme l'expérience dans le domaine de l'évaluation) sont considérées comme qualifiées sur le plan technique. C'est une bonne approche pour les achats ou les services basiques, mais pas pour choisir un service nécessitant un niveau élevé de compétences techniques complexes comme c'est le cas pour les évaluations.
- En additionnant les notes technique et financière pour établir la différence entre l'offre A et l'offre la moins chère acceptable. C'est probablement la meilleure approche car c'est celle qui donne le plus d'importance aux compétences de l'équipe d'évaluation.
- A la note technique multipliée par une note financière pour établir la différence entre l'offre A et l'offre minimale acceptable.

Multiplier la note technique globale par le coût fait du coût un élément prépondérant de la sélection. Dans ce cas, utiliser la première option fait gagner l'offre 2 alors qu'elle n'a obtenu que 73 % des points techniques contre 90 % pour l'offre 1.

Limiter l'impact du coût à certains facteurs est une meilleure approche qui est utilisée par l'UNICEF et les bailleurs scandinaves.

Élément		Offre 1	Offre 2
Pourcentage du nombre maximal de points techniques	A	90 %	73 %
Coût relatif (par rapport au moins cher)	B	1,25	1
Facteur financier (1 / B)	C	0,8	1
Formule de l'UE (A x C)		72 %	73 %
Note technique (80 % by A)	D	72 %	58 %
Note financière (20% x C)	E	16 %	20 %
Formule de l'UNICEF (D + E)		88 %	78 %

## 9.6 Les défis liés au *leadership* et au travail en équipe

Comme tout autre projet, une EAH doit être bien gérée pour réussir. C'est pourquoi Grieg (2011: 3) considère que de bonnes compétences en gestion de projet sont une compétence clé des évaluateurs dans la liste qu'il propose :

1. Compétence technique
2. Fortes compétences en gestion de projet
3. Excellentes capacités de réflexion
4. Excellente capacité à communiquer avec les parties prenantes
5. Flexibilité
6. Tendance à collaborer.

Les trois derniers points sont des compétences interpersonnelles. Pour l'EAH, la compétence technique comprend la connaissance des principes humanitaires et du système humanitaire, ainsi que la capacité à travailler dans des environnements stressants ou peu sûrs.

Il existe potentiellement trois niveaux de gestion : la gestion assurée par le département de l'évaluation de l'organisation commanditaire (le responsable de l'évaluation), la gestion assurée par le responsable du contrat lorsqu'une société de consultants est utilisée, et la gestion du processus et de l'équipe par le chef d'équipe. Chacun de ces niveaux a ses problématiques propres. Cette sous-section se concentre sur la gestion que le chef d'équipe assure au niveau du processus de l'EAH et de l'équipe d'évaluation.

Même lorsque des évaluateurs sont compétents, que les ressources sont adéquates et que les objectifs sont clairs, une bonne planification demeure essentielle pour la réussite d'une évaluation. Presque toutes les EAH rencontrent des contraintes de temps. Se mettre d'accord sur les TdR, recruter une équipe, rédiger le rapport provisoire et le faire circuler, puis intégrer les commentaires des relecteurs prennent presque toujours plus de temps que prévu. Et c'est souvent sur la phase de travail sur le terrain que les retards et dépassements de budget ont le plus d'impact. Une planification minutieuse peut aider à éviter cela.



### Astuce :

Clarifier dès le départ quel appui fournira l'organisation commanditaire à l'équipe d'évaluation. L'équipe aura-t-elle accès aux documents, à un espace de travail au siège et sur le terrain, à de l'aide pour prendre les rendez-vous au siège et sur le terrain, à des services de sécurité, au transport et à un appui pour la réservation du logement sur le terrain ?



L'échelle du travail de l'évaluation détermine la taille de l'équipe. Toutefois, comme nous l'avons mentionné précédemment, les équipes de plus grande taille vont de pair avec des problèmes et des risques qui peuvent dépasser les bénéfices de leur plus vaste champ d'expertise. Les équipes de plus grande taille sont synonymes de plus de travail pour le chef d'équipe, par exemple pour ce qui est de gérer les missions et de rassembler les données. Elles posent également problème dans des environnements peu sûrs. Un logement confortable peut être difficile à trouver dans certains contextes, et les grandes équipes doivent parfois se scinder, ce qui rend difficile les discussions informelles en dehors des heures de travail.

**Astuce :**

Développez des modèles de documents, comme pour les listes de personnes rencontrées, les bibliographies et les itinéraires, dans un format permettant d'obtenir l'information telle que vous la souhaitez. Obtenir des données sous un format standard réduit en effet le travail de conversion et de compilation.

Les équipes d'EAH sont en général constituées via une société de consultants ou une embauche directe. Dans les deux cas, le chef d'équipe peut ne pas avoir travaillé avec l'ensemble des membres de l'équipe par le passé.

**Astuce :**

Si vous travaillez avec certains évaluateurs pour la première fois, et que vous ne connaissez pas leur réputation, organisez le travail sur le terrain de manière à passer une journée ensemble au début de l'évaluation et à pouvoir vous faire une idée de leurs forces et faiblesses, notamment au niveau de leurs biais potentiels.

Même les meilleurs évaluateurs ont des jours « sans » et une bonne performance précédente n'est pas une garantie de bonne performance pour l'évaluation en cours. Un membre de l'équipe peut être affecté par des problèmes personnels, comme la maladie d'un proche. Les membres de l'équipe qui rencontrent des difficultés personnelles peuvent avoir besoin d'un soutien attentif du chef d'équipe. Celui-ci doit néanmoins définir des dates butoir pour les différentes tâches et suivre immédiatement si ces dates sont bien tenues. Cela peut aider à identifier et corriger rapidement tous les problèmes de performance.

Parfois, des problèmes se produisent au sein de l'équipe en raison de conflits de personnalités, de problèmes de performance ou encore de valeurs divergentes. Certains environnements d'EAH peuvent être stressants. Parmi ces problèmes, les questions de performance sont les plus faciles à traiter. Toutefois, le temps que ce problème émerge de façon claire, il peut être déjà trop tard pour écarter la personne identifiée ou la remplacer sans mettre en péril l'évaluation.



**Astuce :**

Si un expert insatisfaisant ne peut être abandonné par l'équipe, mettez-le en binôme avec un membre plus compétent de l'équipe pour aider à minimiser les problèmes de qualité.

Les grandes équipes sont plus susceptibles de connaître des conflits liés aux conclusions, que ce soit à cause de différences de points de vue ou de perceptions du contexte, ou en raison de différences entre les zones que les différents membres de l'équipe ont observées. Voici quelques conseils testés pour gérer les grandes équipes d'évaluation :

- Au début du travail sur le terrain, prévoyez du temps pour que l'équipe apprenne à se connaître mais aussi pour clarifier les rôles et les attentes. Trouvez un moyen pour que chacun puisse partager son domaine d'expertise. Même lorsque le temps manque, il peut être rentable d'investir dans la culture et la communication de l'équipe dès le départ.
- Assurez-vous que tous les membres de l'équipe partagent le même logement afin de permettre des discussions informelles et des relations sociales.
- Utilisez un tableau de synthèse des observations (voir Section 16 : L'analyse) pour garder une trace des éléments de preuve et des premiers résultats. Le fait de partager ces éléments peut permettre de maintenir l'équipe informée et de construire un point de vue commun.
- Si une grande équipe est déployée dans un seul pays, consacrez un peu de temps à une réflexion sur le programme de travail de façon à ce que l'équipe puisse discuter des premiers résultats.
- Planifiez des réunions d'équipe après le travail sur le terrain et avant la finalisation de la première version du rapport dans le but d'échanger sur les données probantes, les résultats, les conclusions et les recommandations.

Parfois, un membre de l'équipe peut être très attentif à un sujet en particulier qui, bien que pertinent, s'avère secondaire par rapport au sujet principal de l'évaluation, ou trop limité, ou encore trop complexe pour être inclus en détail dans le rapport global. Dans ce cas, vous pouvez demander à ce membre de l'équipe d'écrire une annexe sur le sujet.

## 9.7 Gérer les conflits

Les disparités entre l'échelle du travail d'évaluation et les ressources disponibles peuvent mener à un conflit entre le responsable d'évaluation et l'équipe d'évaluation. Un grand nombre de ces conflits proviennent de différences de compréhension. Le rapport de démarrage peut réduire le risque d'incompréhensions durant la phase initiale, mais un conflit peut aussi émerger lors de la rédaction du rapport, entraînant de multiples révisions et une frustration accrue des deux côtés.

**Astuce :**

L'une des façons de minimiser les conflits dans le cadre d'évaluations litigieuses est de faire appel à un groupe de pilotage pour se faire conseiller sur la recevabilité des rapports (même si la décision finale appartient toujours au responsable d'évaluation). Les groupes de pilotage peuvent également contribuer à assurer la qualité de l'ensemble du processus d'évaluation.

Les conflits de personnalité sont une autre source potentielle de conflit. Il peut être utile d'adopter une politique formelle de résolution des conflits spécifiant les différentes mesures qui seront prises en cas de différends et identifiant une personne neutre qui pourrait servir d'arbitre entre les parties si nécessaire.

**Exemple de bonnes pratiques : Développer une politique de résolution des conflits**

La Tsunami Evaluation Coalition a adopté une politique sur la résolution des différends au sein des équipes d'évaluation, entre une équipe et le comité de pilotage, au sein de l'équipe chargée de la synthèse et entre l'équipe chargée de la synthèse et le groupe de direction. Cette politique stipulait que l'auteur principal était responsable de la gestion des relations au sein de l'équipe.

Lorsqu'il y avait un conflit important entre les membres de l'équipe sur une question de fond, ou entre un membre de l'équipe et le chef, que l'auteur principal n'était pas capable de régler, le groupe de direction demandait au Secrétariat d'ALNAP de préparer un rapport présentant les deux versions du problème. Le chef du Secrétariat d'ALNAP pouvait rédiger ce rapport ou embaucher un évaluateur expérimenté pour le faire. Le groupe de direction pouvait ensuite décider de demander à l'auteur principal de mener l'une des actions suivantes :

- Ne présenter qu'une seule des interprétations
- Inclure les deux interprétations dans le rapport
- N'inclure qu'une seule interprétation, mais mentionner qu'elle ne faisait pas l'unanimité.

Si des membres de l'équipe de synthèse n'étaient pas satisfaits de la résolution du problème, ils avaient le droit d'exiger que leur nom soit supprimé du rapport.

**Astuce :**

Les politiques formelles de résolution des conflits sont appropriées pour toute évaluation nécessitant de telles procédures, et pas seulement pour les évaluations de grande échelle.

# 10 / Les méthodes de recherche documentaire

Cette section se focalise sur l'utilisation de la recherche documentaire dans le cadre d'une évaluation, et décrit combien de temps doit être consacré à la revue documentaire de départ. Elle présente également plusieurs manières permettant à l'équipe d'évaluation de présenter de façon succincte les données issues d'une recherche documentaire, pour les briefings et pour les rapports.

Les revues et études utilisent des méthodes de recherche documentaire dans le but de synthétiser les documents.

## Les questions d'évaluation déterminent les méthodes de recherche documentaire

Comme pour les méthodes de terrain, l'ampleur et la profondeur des méthodes de recherche documentaire sont déterminées par les questions d'évaluation. Il n'est possible de répondre à certaines questions d'évaluation que par des méthodes de recherche documentaire. Elles seront même la meilleure manière de répondre pour certaines questions. Par exemple, une question d'évaluation qui cherche à savoir si une intervention atteint sa cible peut trouver sa meilleure réponse dans l'étude de rapports de suivi, puis dans la triangulation des données avec des entretiens sur le terrain.

Voir [Section 6 : Choisir les questions d'évaluation](#) pour plus d'éléments sur les questions d'évaluation.

## 10.1 Qu'est-ce qu'une étude documentaire ?

Une étude documentaire est l'étude d'un ou plusieurs documents. Elle peut intervenir :

- Dans le cadre de [la phase de démarrage](#) pour clarifier le travail de l'évaluation ou répondre à des questions précises
- Dans le cadre de l'exercice de cadrage de l'évaluation pour la préparation des [TdR](#)
- Durant le [travail sur le terrain](#).

Dans certains cas, l'étude documentaire constitue l'évaluation en tant que telle, comme ce fut le cas pour l'étude de la performance du fonds de réponse d'urgence du Consortium des organisations humanitaires britanniques (Stoddard, 2011) ou la revue documentaire des conséquences non intentionnelles de l'aide norvégienne (Wiig et Holm-Hansen, 2014).<sup>3</sup>

Les revues documentaires sont aussi couramment utilisées comme base pour développer des études de capitalisation, comme les études de capitalisation d'ALNAP sur les réponses aux séismes (Cosgrave, 2008), aux catastrophes en milieu urbain (Sanderson *et al.*, 2012) et aux inondations (Cosgrave, 2014). De telles études peuvent s'appuyer sur des évaluations, comme l'étude de la Banque mondiale portant sur des évaluations passées dans le but de nourrir la réponse aux inondations en Afrique de l'Ouest (IEG, 2010) ou sur un ensemble plus large de documents comme la revue de capitalisation du Refugee Studies Centre qui étudiait une décennie de réflexion sur les situations de réfugiés de longue durée (Milner et Loescher, 2011).

La revue documentaire peut être structurée ou non. Les revues documentaires structurées utilisent une structure formelle pour l'analyse des documents, alors que les revues non structurées sont en fait de simples lectures de fond.



**Astuce :**

L'écrit n'est pas automatiquement synonyme de précis ou de fiable. Les données écrites sont souvent considérées comme plus précises ou plus fiables que les données issues d'entretiens. Ce n'est pas nécessairement le cas, et toutes les données provenant de documents ou d'entretiens doivent être testées au moyen d'une triangulation.

Une revue documentaire comprend :

- L'identification des documents. En général, l'organisation commanditaire identifie une première série de documents, mais il revient à l'équipe d'évaluation d'identifier des documentaires complémentaires.
- La lecture (ou autre forme d'analyse) des documents.

La revue documentaire peut être réalisée :

- Par l'équipe d'évaluation, qu'il s'agisse d'une petite tâche de démarrage ou d'une activité majeure de l'évaluation. La revue de la littérature réalisée pour la seconde phase de l'évaluation du lien urgence-réhabilitation-développement (LRRD) suite au séisme et aux tsunamis dans l'océan Indien en est un exemple (Cosgrave *et al.*, 2009).
- Par des consultants engagés spécialement par l'organisation commanditaire. Cela a été fait pour l'évaluation de la thématique LRRD dans le cadre de l'évaluation de la réponse au séisme et aux tsunamis dans l'océan Indien en 2004 (Buchanan-Smith et Fabbri, 2005). Dans cet exemple, les consultants ont identifié et analysé tous les documents.
- Par l'organisation commanditaire, même si c'est rarement le cas sauf lorsque la revue documentaire fait partie de l'exercice de cadrage d'une évaluation. En général, l'implication de l'organisation commanditaire s'arrête à l'identification des documents clés que l'équipe d'évaluation devra prendre en compte.

## 10.2 Pourquoi réaliser une revue documentaire ?

- Dans le cadre d'une évaluation, les revues documentaires sont une activité rentable permettant :
- De s'appuyer sur les connaissances tirées d'évaluations et de recherches précédentes. A titre d'exemple, l'évaluation de Swiss Solidarity portant sur sa réponse aux tsunamis de 2004 dans l'océan Indien a élaboré la première version de son questionnaire d'enquête à partir d'une analyse de précédentes études d'impact (Ferb et Fabbri, 2014: 17).
- De s'appuyer sur les connaissances tirées de documents de suivi de projets.
- A l'équipe d'affiner rapidement sa compréhension du contexte.
- D'identifier les sujets clés potentiels pour le travail sur le terrain à venir. L'étude thématique du PAM sur la nutrition des mères et des enfants a été réalisée en 2002 au moyen d'une revue documentaire qui identifiait les différences clés au sein des opérations du PAM dans différentes régions (Hoogendoorn *et al.*, 2005: 10-11).
- D'identifier des potentiels critères de jugement, sources, et méthodes pour la matrice d'évaluation.

Pour ces diverses raisons, toutes les évaluations devraient comprendre une revue documentaire. Même une évaluation à faible budget a besoin que l'équipe d'évaluateurs analyse au minimum les documents du projet ainsi que toute autre évaluation antérieure jugée pertinente.

Dans l'EAH, les revues documentaires sont particulièrement utiles pour :

- Établir une chronologie des événements passés avec des dates précises. Ces éléments sont essentiels si l'évaluation doit prendre en compte les questions de *timing* ou d'efficacité.
- Montrer comment les priorités de l'intervention ont changé au fil du temps. Les crises humanitaires sont fluides et le contexte peut vite changer. Le renouvellement rapide du personnel clé lors des crises humanitaires signifie que l'équipe de gestion actuelle peut ne pas savoir pourquoi certaines approches spécifiques ont été adoptées. Une revue documentaire peut aider l'équipe d'évaluation à comprendre l'évolution de la réponse.
- Développer une compréhension de points de vue divergents. C'est particulièrement le cas lors d'urgences complexes, car les acteurs humanitaires peuvent adopter celui de la population qu'elles aident. Une revue documentaire peut ainsi aider l'équipe d'évaluation à acquérir une compréhension plus vaste du contexte.
- Identifier les leçons apprises d'opérations précédentes.

## 10.3 Combien de temps doit durer une revue documentaire ?

Le temps nécessaire à la revue documentaire dépend des facteurs suivants :

- La/les question(s) d'évaluation et l'importance des documents seront sûrement une source d'information utile à ce sujet. Si les documents sont censés être une ressource clé pour répondre aux questions, alors il faudra allouer plus de temps à la revue.
- La nature et le périmètre de l'évaluation. Si l'évaluation s'intéresse à une longue période d'intervention, les documents seront sûrement des sources importantes de données et la revue documentaire demandera plus de temps.
- La richesse de l'ensemble des documents disponibles. Des crises récentes peuvent avoir généré peu de documents en dehors des rapports de situation et des diagnostics des besoins. Des crises prolongées peuvent générer des milliers de documents. Si la liste des documents est vaste, alors il faudra allouer plus de temps à la revue.
- La disponibilité de résumés d'ouvrages analytiques. S'ils existent déjà, alors la revue documentaire prendra moins de temps.

## 10.4 Réaliser une recherche documentaire

L'équipe de recherche de l'évaluation du CDC fournit des conseils succincts sur la revue documentaire dans le cadre d'une évaluation (Evaluation Research Team, 2009). Hagen-Zanker et Mallett (2013) donnent des indications sur des revues de littérature rigoureuses dans le domaine du développement international.

### Etape 1 : Identifiez les sources possibles

Les sources des documents dépendent des questions d'évaluation et du temps disponible pour la revue. Les sources des documents peuvent comprendre :

- Les documents clés mentionnés dans les TdR. Ceux-ci peuvent comprendre des documents spécifiquement liés au projet ou au programme ainsi que des documents stratégiques plus larges.
- Les documents propres à l'organisation. Ceux-ci peuvent comprendre des documents clés en lien avec l'évaluation (certaines organisations commanditaires les remettent dans un CD-ROM ou les regroupent dans un dossier en ligne). Les sites Internet ou les intranets des organisations peuvent également être une riche source de diagnostics des besoins, de rapports de situation, de propositions de projet, et de rapports de suivi.
- ReliefWeb, les portails propres à une crise (comme le Portal Haïti d'ALNAP)<sup>4</sup> et les portails thématiques (comme le Urban Humanitarian Response Portal).<sup>5</sup>



- Le Humanitarian Evaluation and Learning Portal (HELP) d'ALNAP.<sup>6</sup>
- Les références et bibliographies existantes.
- Les recherches sur Internet.
- Les recherches dans Google Scholar.
- Les recherches dans les bases de données universitaires.

**Astuce :**

Rédigez une liste des caractéristiques pour les documents que vous recherchez. Vous pouvez décider au départ de vous concentrer sur les évaluations et les études, ou sur les documents postérieurs à une certaine date – cela peut accélérer la revue documentaire.

## Etape 2 : Classez vos documents par catégorie

Il ne faut pas accorder la même valeur ou la même attention à tous les documents. Il peut être utile de classer les documents disponibles en trois groupes :

### Groupe 1

Les documents clés qui cadrent le sujet de l'évaluation sont souvent listés dans les TdR. Ils comprennent les principaux documents de stratégie et de planification du programme et des évaluations passées. En général, on compte entre cinq et vingt documents clés bien que les évaluations au périmètre plus large, couvrant plusieurs pays, secteurs et de longues périodes, puissent en compter davantage.

### Groupe 2

Les documents spécialement consacrés au sujet de l'évaluation, comme les rapports de situation, rapports intermédiaires, les rapports de suivi et les propositions de projet (lorsque l'évaluation concerne un programme plus large). Le nombre de documents varie de vingt à plusieurs milliers, selon le périmètre de l'évaluation.

### Groupe 3

Les documents de fond comprenant la couverture médiatique et les rapports d'autres organisations. Les publications de ReliefWeb sont un exemple de ce type de documents : elles peuvent dépasser les 700 par semaine pour une crise nouvelle, ce qui veut dire qu'il peut y avoir très peu de documents mais aussi des centaines ou même des milliers, selon le périmètre de l'évaluation.



### Etape 3 : Décidez si vous devez adopter une approche structurée

Beaucoup d'évaluations se limitent au premier groupe de documents, et une approche non structurée convient alors. En revanche, une revue documentaire qui comprend des documents issus des deuxième et troisième groupes nécessite une approche structurée, c'est-à-dire :

- Un outil d'observation structurée permettant de consigner les commentaires (voir exemple ci-dessous)
- Utiliser des grilles d'évaluation pour noter certains aspects des documents (voir exemple ci-dessous)
- Indexer et rechercher des documents pour analyser le contenu

### Etape 4 : Faites correspondre la stratégie de la revue documentaire avec le groupe

L'équipe d'évaluation doit lire les documents du premier groupe dans les meilleurs délais. Le deuxième groupe peut par exemple utiliser des grilles d'évaluation et un outil d'enregistrement des observations des lecteurs. Les documents du troisième groupe peuvent être sujets à une analyse de contenu.



**Astuce :**

Utilisez un logiciel de gestion bibliographique, que ce soit des produits commerciaux comme Endnote, initialement gratuits comme Mendeley ou gratuits comme Zotero, qui sont essentiels si vous travaillez sur de nombreux documents. Ils permettent de partager des bibliothèques de référence au sein de l'équipe et garantissent que les documents seront cités de façon harmonisée.



**Astuce :**

Utilisez une application cloud pour partager les documents (Dropbox, Box, One Drive ou Google Drive) afin de conserver une copie des documents et de la rendre disponible à tous les membres de l'équipe.

## 10.5 Les outils de la recherche documentaire

En matière de recherche documentaire, les formulaires qui permettent d'enregistrer les commentaires de façon structurée font partie des outils clés. Ils peuvent être combinés à l'emploi de grilles d'évaluation permettant de noter les documents sur des aspects particuliers.

### Les grilles d'évaluation

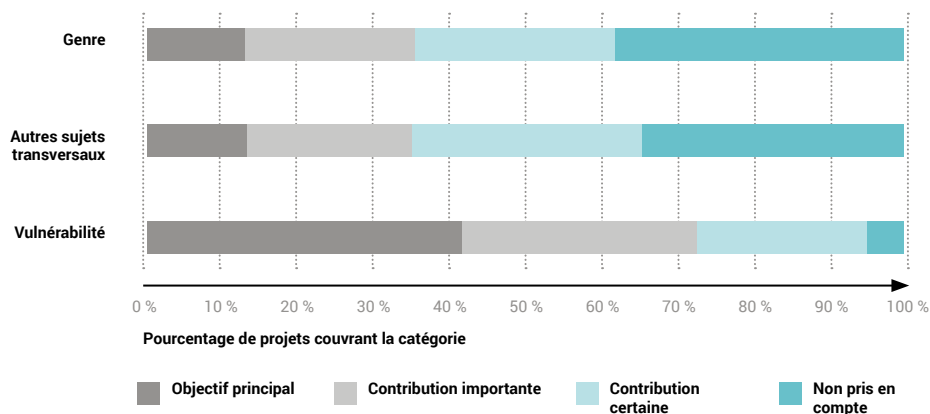


#### Définition : Grille d'évaluation

Une grille d'évaluation est un outil de notation utilisé pour évaluer un document à partir d'une série de critères de façon harmonisée.

Les notes décernées peuvent être utilisées pour classer les documents ou pour comparer les changements de cap au fil du temps. Par exemple, l'évaluation de cinq ans du Central Emergency Response Fund (CERF) a utilisé des grilles d'évaluation pour le genre, la vulnérabilité et d'autres sujets transversaux pour conclure que, tandis que 95 % des projets prévoyaient de contribuer au moins à la réduction de la vulnérabilité, moins de deux tiers prévoyaient de contribuer à la promotion de l'égalité des sexes et d'autres sujets transversaux (Channel Research, 2011).

Schéma 10.1 : Genre, vulnérabilité et sujets transversaux



Les notations genre, vulnérabilité et sujets transversaux appliqués à un échantillon aléatoire de 258 projets financés par le CERF et choisis dans 16 pays d'étude. 41 projets ont été exclus faute d'informations suffisantes ou parce qu'ils correspondaient à des services communs.

La notation genre (IASC, 2010) est un exemple de grille d'évaluation. Il est plus fréquemment utilisé pour évaluer la conception des projets.

Niveau	Description
0	Le genre n'est traité dans aucune composante du projet.
1	Le projet est conçu pour contribuer de façon limitée à l'égalité des sexes. Les dimensions liées au genre sont prises en compte de façon significative uniquement dans une ou deux des trois composantes essentielles : diagnostic des besoins, activités et résultats.
2a	<b>Approche intégrée du genre</b> : le projet est conçu pour contribuer de façon significative à l'égalité des sexes. Les besoins différents des femmes/filles et des hommes/garçons ont été analysés et bien pris en compte dans les trois composantes essentielles : diagnostic des besoins, activités et résultats.
2b	<b>Actions ciblées</b> : l'objectif principal du projet est de favoriser l'égalité des sexes. L'ensemble du projet cible soit (a) les femmes ou les hommes, les filles ou les garçons qui ont des besoins spécifiques ou souffrent de discrimination, soit (b) focalise toutes les activités sur la création de services propres au genre ou de relations plus égalitaires entre femmes et hommes.

On note un intérêt croissant pour l'utilisation de grilles d'évaluation, et pas seulement pour la revue documentaire (Oakden, 2013a ; 2013b ; Rogers, 2013). Voir également l'utilisation de mini-grilles d'évaluation dans les ateliers (Davidson, 2014).

Des grilles d'évaluation similaires peuvent être développées pour tout aspect pertinent. Une grille d'évaluation peut impliquer une échelle plus complexe ou se contenter d'utiliser une échelle de quatre points pour signifier le niveau d'attention vis-à-vis de chaque sujet, comme le handicap, la lutte contre la corruption, la RRC ou les droits des enfants. Une simple échelle à quatre points peut diviser les documents de la façon suivante :

- Ceux qui ne traitent pas du tout le sujet.
- Ceux qui traitent le sujet de façon limitée. La grille d'évaluation doit expliciter ce qui est vu comme limité.
- Ceux qui traitent le sujet de façon significative. La grille d'évaluation doit expliciter ce qui est considéré comme significatif.
- Ceux qui se focalisent principalement sur le sujet.

Les grilles d'évaluation sont utiles pour garantir :

- La cohérence entre les différents évaluateurs.
- La cohérence au fil du temps d'un même évaluateur.

Il est utile qu'une deuxième personne réexamine une sélection aléatoire de documents notés afin de rechercher des biais ou des notations incohérentes.

## Les formes structurées de recherche documentaire

Vous trouverez ci-dessous un exemple de formulaire simple où l'évaluateur peut noter différents aspects d'un document en vue d'un suivi ultérieur. Par exemple, si l'intérêt se porte sur la coordination intra- et intersectorielle, il est possible d'examiner les documents avec une première lecture rapide et de noter les éléments intéressants.

Pour structurer une recherche documentaire, une autre option consiste à utiliser le tableau de synthèse des observations décrit dans la [Section 16 : L'analyse](#). Les formulaires structurés de recherche documentaire peuvent également être utilisés pour consigner les notations des grilles d'évaluation sur différents aspects.

Doc	Page	Coordination sectorielle	Coordination intersectorielle
1	6	« Participation régulière à des réunions Cluster. »	
	9	« Nous mettons immédiatement en œuvre la décision du Cluster WASH de chlorer à la source toute l'eau transportée en citerne et d'analyser les niveaux de chlore à la sortie. »	
	19	« 87 % des citernes répondent aux normes du Cluster. Celles qui ne satisfont jamais ces exigences ont été retirées du contrat et les autres membres du Cluster ont été informés des numéros de plaque. »	« L'équipe Eau et assainissement a rencontré l'équipe centrale du Cluster Logistique pour évoquer la question du traitement des déchets solides. »

## L'analyse de contenu

L'analyse de contenu est une manière utile de gérer un grand nombre de documents. La plupart des outils les plus utilisés en matière d'analyse des données qualitatives offrent la possibilité de faire une analyse de contenu mais ils ne sont pas toujours faciles à appliquer. Une alternative moins coûteuse et plus simple consiste à utiliser un outil d'indexation comme dtSearch<sup>7</sup> pour rechercher des mots dans les documents-clés ou réaliser d'autres types d'analyse de contenu.



### Définition : Analyse de contenu

L'analyse de contenu est une analyse de l'information textuelle d'une manière standardisée permettant aux évaluateurs de procéder à des déductions au sujet de l'information.

L'analyse de contenu prend en général la forme d'un codage des documents dans le but de réduire le texte à une série de catégories. Une forme d'analyse de contenu utilisée dans un certain nombre d'évaluations, y compris dans l'évaluation de cinq ans du CERF (Channel Research, 2011), est l'analyse des mots-clés.



### Définition : Analyse des mots-clés

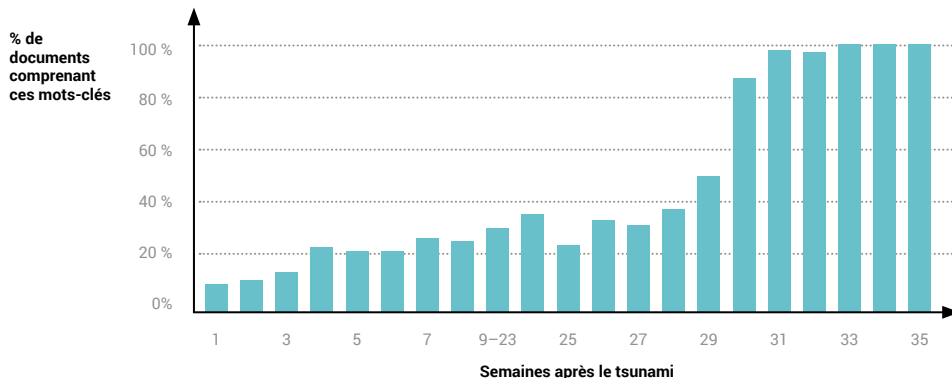
L'analyse des mots-clés est une forme d'analyse de contenu qui examine la fréquence de l'occurrence des catégories pour mettre en évidence les tendances au fil du temps dans une seule série de documents ou pour en comparer deux.

Une liste de mots-clés peut présenter jusqu'à 500 termes. Certaines catégories seront représentées par plusieurs mots-clés. La catégorie Eau, Assainissement et Hygiène, comprend par exemple les mots-clés suivants : puits de forage, chlore, chloration, défécation, matières fécales, hygiène, latrines, matières de vidange, assainissement, savon, source, toilettes, WASH, eau, approvisionnement en eau, watsan, puits.

Les autres formes d'analyse de contenu sont rarement utilisées dans l'EAH. Une exception eut lieu lors de l'évaluation d'impact de l'aide à l'intégration des rapatriés au Burundi où des entretiens retranscrits ont été sujets à une analyse de contenu (Telyukov *et al.*, 2009).

Trois guides sont particulièrement utiles en matière d'analyse textuelle : Krippendorff (2004), document de référence ; Benini (2009), un bref guide destiné aux acteurs humanitaires et du développement qui décrit comment utiliser trois outils d'analyse textuelle ; et le guide du US General Accounting Office (GAO, 1996). L'analyse de contenu peut être utilisée pour montrer l'émergence et l'évolution des différents concepts au fil du temps. L'exemple qui suit montre la prévalence de termes liés à la formation dans la série de documents de ReliefWeb portant sur l'évaluation de la Tsunami Evaluation Coalition suite au séisme et aux tsunamis qui se sont produits en 2004 dans l'océan Indien. On peut ainsi voir qu'après trente jours presque tous les documents renvoyaient d'une manière ou d'une autre à la formation, alors qu'au départ on y accordait peu d'attention.

Schéma 10.2 : Prévalence des mots-clés liés à la formation dans 14.560 documents sur le tsunami



## 10.6 Les produits issus de la revue documentaire

Les conclusions des revues documentaires réalisées par l'équipe d'évaluation peuvent être intégrées dans des réunions de briefing, dans le rapport de démarrage, et dans le principal rapport d'évaluation.

Un défi particulier consiste à condenser une grande quantité de données issues d'une revue documentaire dans un format facilement accessible. Cela peut être fait sous la forme de tableaux et de graphiques. Synthétiser les données de cette manière peut également aider à informer l'ensemble de l'équipe d'évaluation.

### Les chronologies

Les chronologies sont très utiles dans l'EAH car elles peuvent aider à traiter la question du séquençage de l'assistance.

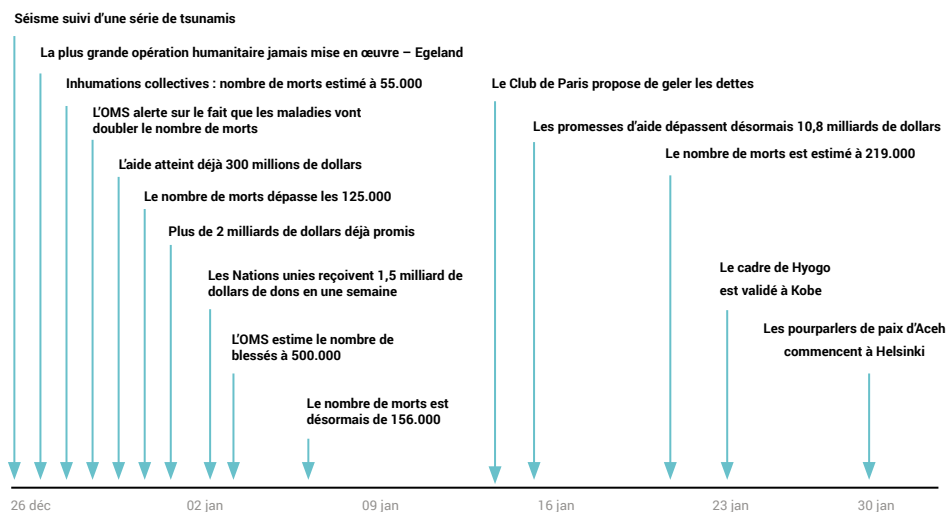


#### Astuce :

Commencez à rassembler dès que possible les données de la chronologie. Débuter tôt la chronologie peut aider à contextualiser d'autres lectures lorsque vous commencez la revue documentaire.

Les chronologies peuvent être présentées sous la forme d'un tableau ou d'un graphique. La chronologie partielle qui suit illustre le déclenchement de la réponse aux tsunamis qui se sont produits en 2004 dans l'océan Indien. Elle indique que des financements ont été engagés bien avant les premiers diagnostics détaillés des besoins.

**Schéma 10.3 :** Chronologie partielle du séisme et des tsunamis du 26 décembre 2004 (Asie)



Les chronologies peuvent être élaborées à partir de documents datés de la série de documents.



**Astuce :**

Marquez les données des documents en ajoutant un préfixe à tous les noms de fichiers de la série de données sous la forme YYYYmmDD. Cela permet de trier rapidement les documents et de vérifier quels documents ont été publiés à quelle date.

Le logiciel d'indexation ou d'analyse de contenu, comme le logiciel dtSearch déjà mentionné, peut avoir la capacité de rechercher des dates à l'intérieur des documents, ce qui peut vous faire gagner du temps au moment d'élaborer une chronologie.



**Astuce :**

Impliquez l'ensemble de l'équipe dans l'élaboration de la chronologie parce qu'elle permet à chacun d'acquérir une bonne compréhension de la manière dont la réponse a évolué.

### Autres présentations de données par ordre chronologique

Les chronologies ne sont qu'une seule forme de données d'ordre chronologique qui peuvent aider l'équipe à comprendre comment une crise s'est développée. Il est par exemple possible de mettre les événements en perspective avec une courbe épidémique, comme ce fut le cas pour le financement du CERF et l'épidémie de dengue au Cap-Vert en 2009.

**Schéma 10.4 :** Frise chronologique de l'épidémie de dengue au Cap-Vert et du processus de demande du CERF

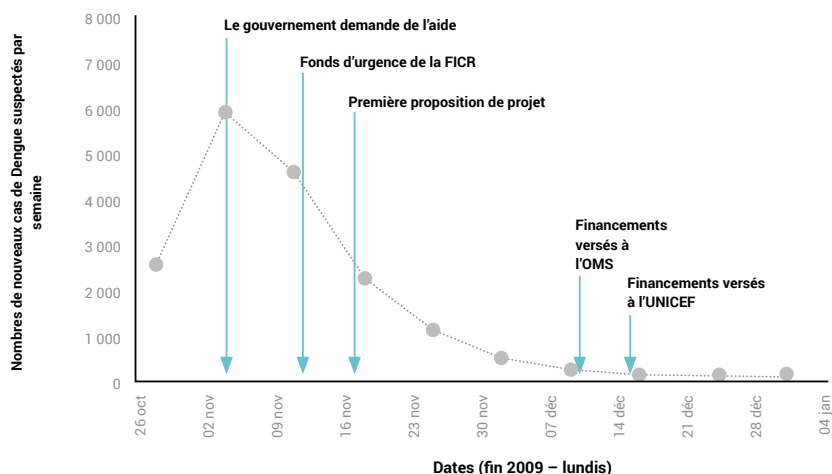
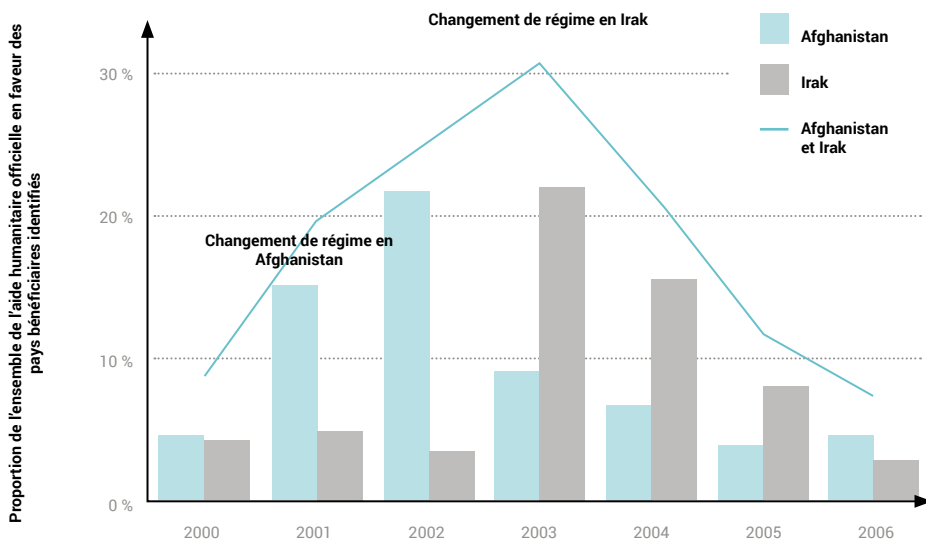


Schéma 10.5 : L'aide humanitaire pour l'Afghanistan et l'Irak (2000-2006)



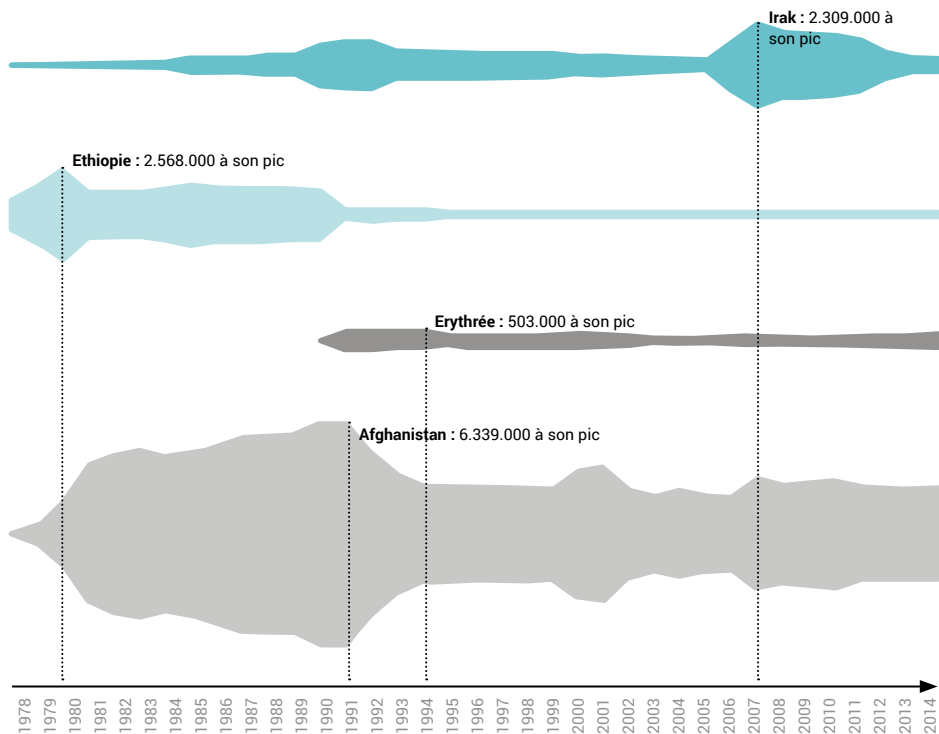
Source : OCDE-CAD, tableau 2a, 10 avril 2008. Deux tiers de l'aide humanitaire sont allés à des pays bénéficiaires identifiés pendant cette période

Les données chronologiques sont également utiles pour les évaluations de politique. Le Schéma 10.5 montre comment les changements de régime en Afghanistan et en Irak sont allés de pair avec une forte augmentation de l'aide humanitaire officielle.

Les flux de réfugiés se prêtent naturellement à une présentation sous forme graphique. Le Schéma 10.6, réalisé pour une étude sur les déplacements prolongés (Crawford *et al.*, 2015), montre l'évolution du nombre de dossiers de réfugiés dans dix pays de 1978 à 2014. La largeur de chaque trace est proportionnelle au nombre de réfugiés à la période indiquée. Ce graphique aide à illustrer l'argument selon lequel il n'existe pas de modèle dominant en matière de déplacement de longue durée, mais que chaque crise suit son propre modèle.



Schéma 10.6 : Sélection de crises de réfugiés générant plus de 400.000 réfugiés entre 1978 et 2014



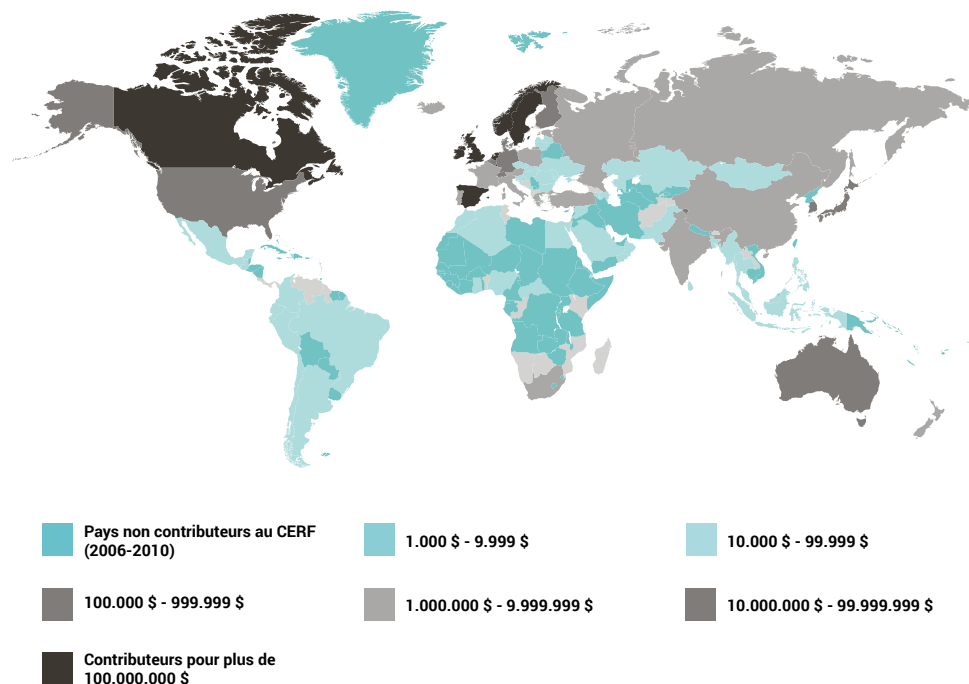
Deux tiers de l'aide humanitaire sont allés à des pays bénéficiaires identifiés pendant cette période Based on UNHCR Data.

## Les données géographiques

Les données géographiques peuvent être synthétisées en utilisant des logiciels graphiques (SIG) qui permettent de faire ressortir des sujets clés. Il existe des solutions SIG peu onéreuses et libres (*open-source*) mais elles ne sont pas toujours faciles à maîtriser. La carte ci-dessous montre quels pays ont contribué au CERF. Durant ses cinq premières années, les couleurs les plus sombres correspondant aux contributions les plus importantes. Elle indique clairement que le CERF bénéficie d'un large soutien, y compris des pays en voie de développement.

A nouveau, le logiciel d'analyse de contenu peut faire ressortir dans quelle mesure des noms géographiques sont utilisés au sein d'une série définie de données et cela peut être utilisé pour montrer leur orientation géographique.

Schéma 10.7 : Contributions par pays au CERF (2006-2010)

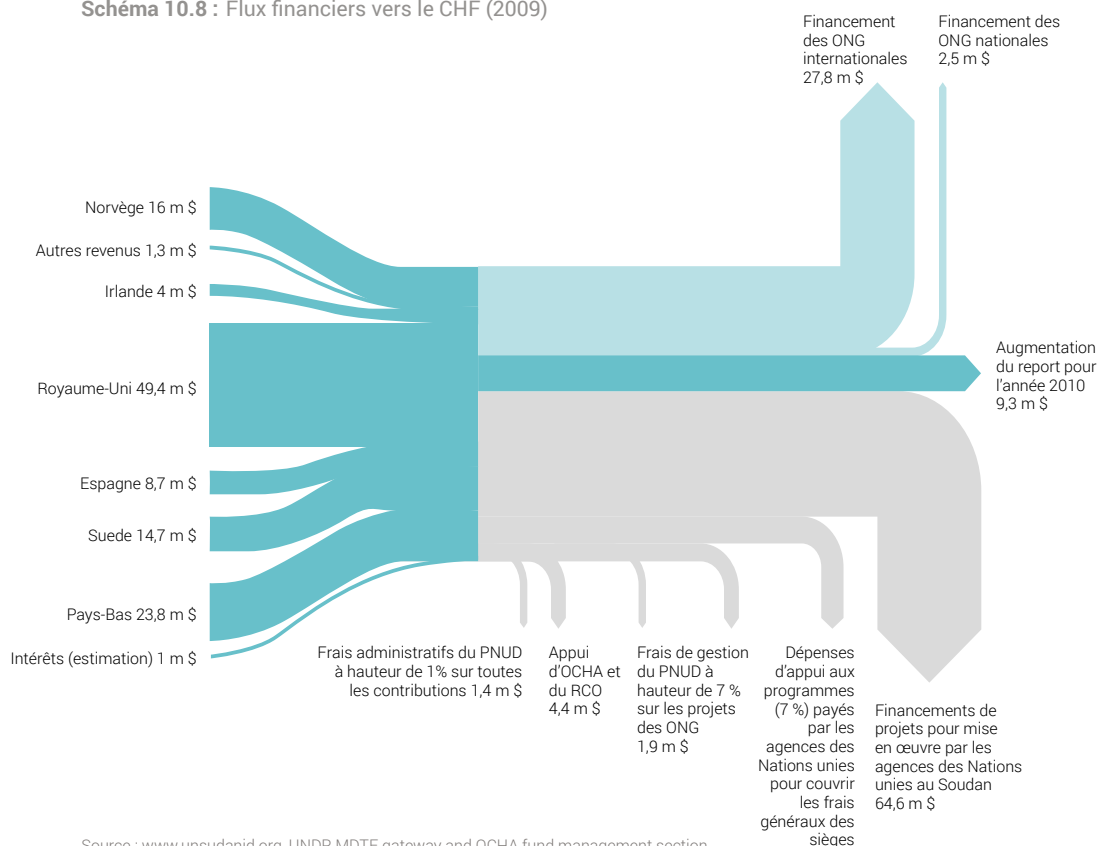


## L'analyse des flux

Il peut être plus facile de comprendre les flux complexes s'ils sont présentés sous la forme d'un diagramme Sankey. Le Schéma 10.8, tiré de l'évaluation du Fonds humanitaire commun (Soudan, 2009), montrait que le financement relativement faible était dirigé vers les ONG nationales, et qu'un montant considérable était consacré aux frais administratifs.

La Section 16 abordera la question de l'analyse numérique.

**Schéma 10.8 :** Flux financiers vers le CHF (2009)



Source : [www.unsudanid.org](http://www.unsudanid.org), UNDP MDTF gateway and OCHA fund management section

## Notes de fin de texte

### 9 / Planifier et gérer votre évaluation

1. Jane Davidson (2010a) inclut le fait de « choisir le prestataire dont le coût journalier est le moindre » dans sa liste ironique des *9 neuf règles pour commanditer une évaluation qui gaspille de l'argent (9 golden rules for commissioning a waste-of-money evaluation)*.
2. Pour certaines agences des Nations unies, il est par exemple plus rapide de recruter directement des évaluateurs que de travailler avec une société de consultants.

### 10 / Les méthodes de recherche documentaire

3. Cette étude a conclu que toutes les évaluations norvégiennes échantillonnées d'aide humanitaire avaient constaté des conséquences non intentionnelles.
4. Voir [www.alnap.org/current/haitilearningportal](http://www.alnap.org/current/haitilearningportal).
5. Voir [www.urban-response.org](http://www.urban-response.org).
6. Voir [www.alnap.org/resources](http://www.alnap.org/resources).
7. Voir <http://dtsearch.com>.

8 / La phase de démarrage  
135 - 146

9 / Planifier et gérer votre évaluation  
147 - 171

10 / Les méthodes de recherche documentaire  
172 - 187

## Notes

## Notes

# Réaliser l'évaluation



# 11 / Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation

Cette section s'intéresse à la manière dont il faut structurer l'évaluation pour répondre à la question d'évaluation, ce qui peut en retour aider les évaluateurs et les responsables d'évaluation à envisager les possibles conceptions d'évaluation.

La conception de l'évaluation est la logique générale selon laquelle vous organisez votre recherche pour répondre à une question d'évaluation. Certaines conceptions d'évaluation répondent mieux que d'autres à des types spécifiques de question. L'objectif de cette section est de vous aider à estimer quelle(s) conception(s) d'évaluation permettra/ permettront à l'équipe d'évaluation de répondre au mieux à la/ aux question(s) d'évaluation étant donné la nature du programme, le contexte de sa mise en œuvre, et les contraintes de l'évaluation, notamment l'accès, le budget, les données, et d'autres facteurs.

En règle générale, une recherche classique a une seule hypothèse ou question à examiner, ou alors un groupe d'hypothèses ou de questions étroitement liées. Toutefois, les évaluations comprennent souvent une gamme de questions disparates et cela peut signifier qu'il faut utiliser plusieurs types de conception d'évaluation ou qu'il faut utiliser des conceptions qui permettent de répondre à toutes les questions mais qui ne sont pas idéales pour certaines d'entre elles.



## **A garder à l'esprit**

Il est important de ne pas confondre conceptions et méthodes. La conception renvoie à la structuration de la collecte et de l'analyse des données, alors que la méthode renvoie à la manière dont les données sont rassemblées.

Comme expliqué dans la [Section 6 : Choisir les questions d'évaluation](#), les questions d'évaluation déterminent la conception de l'évaluation, la collecte des données et les méthodes d'analyse, ainsi que les approches par échantillonnage.

Cette section aborde également le problème des biais, ces derniers pouvant varier selon la conception, les méthodes et les approches par échantillonnage de l'évaluation.



## 11.1 Qualitatives et quantitatives

Certaines conceptions d'évaluation (comme les études de cas) sont couramment associées à des approches qualitatives et d'autres (comme les expérimentations) à des approches quantitatives.



### Pour approfondir : Le problème de l'opposition qualitatif vs. quantitatif

En principe, les méthodes quantitatives collectent des données numériques et les méthodes qualitatives des données non-numériques. Toutefois, dans la pratique, ces étiquettes sont beaucoup plus complexes que cela. Goetz et Mahoney (2012) les décrivent comme représentant deux différentes cultures de recherche. Mahoney et Goetz (2006: 245) remarquent que « les étiquettes quantitative et qualitative sont peu pertinentes pour saisir les réelles différences entre les traditions. L'analyse quantitative implique de façon inhérente l'utilisation des nombres, mais toutes les analyses statistiques s'appuient fortement sur les mots pour l'interprétation. Les études qualitatives ont souvent recours à des données numériques ; et de nombreuses techniques qualitatives nécessitent en fait de l'information quantitative.

Les faiblesses de ces étiquettes ont amené à en proposer plusieurs autres, notamment « small-n », explicative et fondée sur des cas, pour la recherche qualitative et « large-n », statistique, estimative et étude de population pour la recherche quantitative. Mahoney et Goetz (2006: 245) listent ainsi dix points de divergence entre elles. Patton (2014, schéma 3.1) utilise dix facteurs pour distinguer les approches quantitatives, qualitatives et combinées.

Ce débat peut sembler de nature académique mais il est important pour les évaluateurs humanitaires parce que les deux cultures présentent des théories du savoir qui sont très différentes. Dans le cas de la recherche quantitative, le savoir tire son origine des méthodes expérimentales, comme dans les sciences physiques. Ce modèle de développement du savoir est appelé positivisme. Dans le cas de la recherche qualitative, il existe plusieurs théories différentes de la formation du savoir, notamment l'ethnographie et la théorie ancrée (« grounded theory »). Patton (2014, ch. 3) liste ainsi 16 différentes approches théoriques utilisées dans le domaine de la recherche qualitative.

Pour les évaluateurs, le problème est le suivant : ces dernières années, une tendance a considéré que seule une connaissance issue de moyens expérimentaux était rigoureuse. L'Association américaine d'évaluation (2003) et la Société européenne d'évaluation (2007) ont toutes deux critiqué cette approche. Knox-Clarke et Darcy (2014) affirment que l'utilisation de méthodes qualitatives ou quantitatives n'aboutit pas de façon automatique à des éléments de preuve plus ou moins probants.



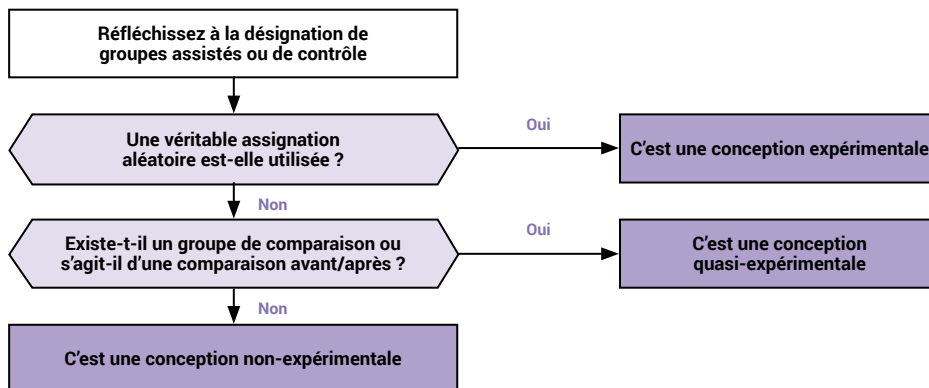
Un autre problème pour les évaluateurs se trouve dans le fait que les organisations commanditaires privilégient les expérimentations ou appliquent des mesures adaptées à un type de recherche pour un autre type de recherche. La recherche qualitative ne peut pas être évaluée en appliquant sans aucune critique la terminologie et les concepts utilisés pour évaluer la recherche quantitative (Anastas, 2004: 57). Un document de l'ACAPS distingue les indicateurs de recherche quantitative et qualitative de la façon suivante : *validité interne/précision, validité externe/« caractère généralisable », fiabilité/cohérence/précision, et objectivité* pour la recherche quantitative contre *crédibilité, transférabilité, fiabilité, et corroboration* pour la recherche qualitative (ACAPS, 2013).

## 11.2 Les familles de conception

Il existe trois grandes familles de conception de recherche (voir exemples ci-dessous):

- Les conceptions non-expérimentales sans comparaison ni groupes de contrôle. C'est une conception non-expérimentale (la plus courante dans l'EAH)
- Les conceptions expérimentales où la désignation des groupes assistés ou de contrôle est faite avant le début de l'intervention.
- Les conceptions quasi-expérimentales où des comparaisons sont faites soit au sein du groupe assisté au fil du temps, soit entre le groupe assisté et un groupe de comparaison sélectionné après le début de l'intervention.

La catégorie dans laquelle s'inscrit une conception d'évaluation dépend de l'utilisation ou non d'un groupe de contrôle ou de comparaison. Les comparaisons peuvent se faire au sein de groupes ou au fil du temps pour le même groupe. The following flowchart summarises the distinctions between the families of EHA design.



Si une conception de recherche n'utilise ni groupe de contrôle ni groupe de comparaison, il s'agit d'une conception non-expérimentale, et c'est le type le plus courant dans l'EAH.

Une seule évaluation peut nécessiter différentes conceptions consécutivement de manière à répondre aux questions de façon adéquate. Cela peut être le cas avec des conceptions combinées qui peuvent comprendre des intervalles de méthodes quantitatives et qualitatives utilisant différentes conceptions.

Les TdR de l'importante évaluation d'impact de Swiss Solidarity portant sur les dix années postérieures au séisme et aux tsunamis dans l'océan Indien demandaient à l'origine qu'une enquête quantitative soit utilisée avant toute autre méthode qualitative. Ferf et Fabbri (2014: 16) ont prôné une inversion de cet ordre afin de « mettre en place des entretiens aussi ouverts que possible, évitant des attentes préconçues au niveau des résultats, et favorisant des opportunités de saisir autant que possible les imprévus. Par la suite, la recherche s'est resserrée sur les questions identifiées comme les plus pertinentes pour le rétablissement des bénéficiaires. »

Cette option peut toujours être améliorée dans un deuxième temps avec un examen non-expérimental de suivi des sujets identifiés dans l'enquête. Ce modèle, qui comprend une étude exploratoire non-expérimentale permettant de définir l'angle de l'enquête quasi-expérimentale, est suivi par un examen approfondi non-expérimental final des sujets émergents, est le cycle de modèle de recherche proposé par Remler et Ryzin (2015), lequel est semblable aux exemples de méthodes combinées dites « en sandwich » que donne Bamberger (2012).



## Les conceptions non-expérimentales

### Définition : Les conceptions non-expérimentales

Les conceptions non-expérimentales sont des conceptions qui n'ont recours à aucune comparaison, que ce soit entre populations assistées et populations non-assistées, ou au fil du temps au sein des populations qui ont reçu une assistance.

La plupart des EAH utilisent des conceptions non-expérimentales ponctuelles, notamment parce que :

- Elles sont les moins exigeantes, au niveau des conditions qui doivent être remplies
- Elles sont les plus flexibles
- Elles sont relativement peu onéreuses
- Elles sont capables de répondre à tout type de questions d'évaluation
- Elles sont bien adaptées aux compétences des évaluateurs humanitaires, étant généralement développées en réalisant les diagnostics des besoins et la planification des programmes.

Les conceptions non-expérimentales peuvent comprendre :

- Les études de cas
- Les études de processus
- Les études de résultats
- Les conceptions participatives.

L'évaluation à base théorique est parfois présentée comme un type de conception non-expérimentale, même si la question est plus complexe que cela. Une évaluation à base théorique peut utiliser diverses conceptions pour tester la théorie de changement (TdC) implicite, notamment les conceptions expérimentales, quasi-expérimentales et non-expérimentales.

Les études de cas, dans lesquelles l'évaluation examine une série de différentes unités d'analyse pour tirer des conclusions générales sur l'intervention, sont probablement la conception la plus courante dans l'EAH. Les véritables conceptions participatives (où les parties prenantes définissent l'évaluation) ne sont pas utilisées dans l'EAH. Voir Section 14 : Impliquer la population affectée dans votre évaluation. Les évaluateurs humanitaires doivent dépasser l'utilisation des seules conceptions non-expérimentales parce que :

- D'autres conceptions sont plus à même de répondre à certains types de questions d'évaluation
- Certains bailleurs exercent des pressions pour que les évaluateurs se rapprochent de conceptions qui sont parfois vues comme plus rigoureuses
- D'autres conceptions peuvent être à même de répondre à certaines questions d'évaluation pour un coût inférieur à celui des conceptions non-expérimentales (cela peut être particulièrement vrai pour les conceptions quasi-expérimentales qui utilisent des données secondaires).

La sous-section sur le choix de votre conception à la page 220 indique quand certaines conceptions sont les plus utiles.

## Les conceptions quasi-expérimentales



### Définition : Les conceptions quasi-expérimentales

Les conceptions quasi-expérimentales sont des conceptions qui utilisent une comparaison lorsque le groupe de comparaison n'est pas choisi de manière aléatoire.

Les quasi-expérimentations couvrent une large gamme de conceptions, dont :

- Les conceptions où le groupe qui reçoit l'assistance est évalué par rapport à un groupe de comparaison (à un moment précis ou sur une période déterminée).
- Les conceptions où un même groupe est comparé dans la durée.

Diverses conceptions quasi-expérimentales sont présentées dans la sous-section « Choisir une conception ». La principale force de ces conceptions est de pouvoir fournir des données probantes rigoureuses pour déterminer si une intervention a été efficace, tout en évitant les problèmes éthiques des conceptions expérimentales.

### L'utilisation de groupes de comparaison

L'utilisation de groupes de comparaison aide à réduire le risque que tout changement constaté dans le groupe recevant l'assistance ne soit dû à des changements de fond plus larges, plutôt qu'à l'aide fournie. Les conceptions utilisant des groupes de comparaison sont plus robustes que les autres.

Toutefois, les groupes de comparaison des contextes humanitaires sont facilement sujets à la contamination expérimentale en raison du grand nombre d'acteurs présents et des autres réseaux de soutien qui opèrent (notamment les familles).

La contamination expérimentale se produit lorsqu'il est impossible de contrôler tous les aspects de l'assistance au niveau des groupes assistés et des groupes de contrôle ou de comparaison.

La contamination des expérimentations et des quasi-expérimentations en contextes humanitaires peut se produire de nombreuses manières, y compris lorsque :

- D'autres organisations ou membres de la famille fournissent une assistance aux membres du groupe de contrôle qui reçoit une aide
- Les membres du groupe de contrôle apprennent du groupe recevant l'aide (par exemple, à propos des bénéfiques liés au fait de se laver les mains avant de toucher la nourriture) et adoptent cette meilleure pratique.

### Quelques exemples des problèmes liés aux groupes de comparaison

Une évaluation des transferts monétaires destinés au retour des réfugiés au Burundi avait prévu de comparer les bénéficiaires et les non-bénéficiaires, mais elle a vite remarqué que les deux groupes ne pouvaient pas être comparés puisqu'ils avaient effectué leur retour à des périodes différentes (Haver et Tennant, 2009).

Une étude portant sur les enfants soldats au Mozambique (*The Mozambique Child Soldier Study*) avait prévu d'utiliser un groupe de comparaison mais l'équipe s'est rendue compte que beaucoup de ceux qu'ils avaient choisis pour des entretiens ne répondraient pas pleinement à leurs questions (Boothby *et al.*, 2006: 106). L'*Inter-Agency Guide to the Evaluation of Psychosocial Programming in Humanitarian Crises* constate que cela a obligé l'étude à utiliser des « normes locales » comme éléments de comparaison plutôt que d'autres enfants soldats (Ager *et al.*, 135).

Une évaluation d'impact de l'aide alimentaire en faveur des réfugiés au Bangladesh a remarqué que la différence entre les groupes assistés et de comparaison menaçait la validité interne de l'étude (Nielsen *et al.*, 2012).

L'étude Milk Matters en Ethiopie a utilisé un groupe de comparaison qui était déjà très différent du groupe assisté, y compris avant le début du traitement (Sadler *et al.*, 2012). Lorsqu'il existe une forte différence avant l'intervention entre les deux groupes, il est impossible d'attribuer toute différence ultérieure à l'intervention.

Les groupes de comparaison peuvent être mis en place grâce à différentes méthodes. La plus rigoureuse est probablement l'appariement par score de propension au moyen de méthodes statistiques.



#### **Définition : L'appariement par score de propension**

L'appariement par score de propension est une technique d'appariement statistique qui cherche à associer le groupe de comparaison avec le groupe de contrôle en choisissant un qui ait la même probabilité de recevoir une aide selon ses caractéristiques.

Caliendo et Kopeinig (2005) donnent des indications sur l'utilisation de l'appariement par score de propension dans le cadre d'une évaluation. Certaines difficultés liées à son application ont émergé durant l'étude de l'alimentation complémentaire de masse au Niger (Grellety *et al.*, 2012).

Il est également possible de choisir le groupe de comparaison en s'appuyant sur l'avis de l'expert.

## Les conceptions expérimentales

Les conceptions expérimentales sont les plus exigeantes et de nombreuses conditions doivent être réunies pour les appliquer avec succès. Certaines personnes les considèrent comme la conception la plus rigoureuse.



### Définition : Les conceptions expérimentales


Les conceptions expérimentales sont celles où les unités d'analyse sont déterminées de façon aléatoire au groupe recevant l'assistance ou au groupe de contrôle. Chaque élément (par exemple, une personne, une famille ou une communauté) a une probabilité égale d'être désigné groupe assisté ou groupe de contrôle.

L'assignation aléatoire de l'assistance pose des problèmes éthiques dans le cadre de l'EAH parce que l'aide doit être fournie sur la base des besoins plutôt que d'être laissée au hasard.


Les essais contrôlés randomisés (RCT sous son acronyme anglais) sont la conception expérimentale la plus courante. Là où des RCT ont été réalisés dans des contextes d'EAH, ils ont parfois été menés entre différentes formes d'assistance (espèces/ coupons ou nourriture, comme pour Sandström et Tchatchua, 2010) plutôt qu'entre assistance et non-assistance, ou en phase de relèvement comme l'évaluation de la reconstruction communautaire dans le Comté de Lofa (Fearon *et al.*, 2008). Des études comprenant des RCT ont été réalisées en santé dans des contextes humanitaires, comme pour celle portant sur l'impact de l'alimentation complémentaire avec probiotiques et prébiotiques au Malawi (Kerac *et al.*, 2009) ou sur les bâches en plastique traitées à l'insecticide en Sierra Leone (Burns *et al.*, 2012).

## 11.3 Choisir une conception

Idéalement, la conception est seulement déterminée par les questions d'évaluation, et aucune conception d'évaluation n'est parfaite. Les contraintes imposées par le timing, le budget, la disponibilité des données et autres problèmes limitent les options. Les options retenues, et les raisons de ces choix, doivent être notées dans les rapports de démarrage et final. En effet, il peut être nécessaire de s'écarter de la conception prévue sur le terrain. Cette section décrit les points clés de plusieurs conceptions. Les parenthèses indiquent le type de conception.

Etude de cas (non-expérimentale)	
 <b>Définition :</b> les études de cas sont une description et une analyse intensive d'un ou plusieurs cas (qui peuvent aller des individus aux Etats) permettant de tirer des conclusions générales sur l'intervention.	
<b>Utilisation dans l'EAH</b>	
Les études de cas sont couramment utilisées, en général par rapport aux familles et aux communautés. Les études de cas pays peuvent être utilisées pour des évaluations de grande échelle.	
<b>Conseils</b>	
Yin (2003a ; 2003b) donne des indications sur l'utilisation des études de cas.	
Points forts	Points faibles
<ul style="list-style-type: none"> <li>• Pertinente pour répondre à des questions cherchant à savoir pourquoi les gens ont agi de telle ou telle manière</li> <li>• Pertinente pour étudier des conditions rares ou des interventions complexes</li> <li>• Pertinente pour tester une théorie</li> <li>• Fournit des données riches sur les cas étudiés</li> <li>• Pertinente pour élaborer une théorie</li> <li>• Bien adaptée à l'EAH</li> </ul>	<ul style="list-style-type: none"> <li>• Peut être difficile à généraliser</li> <li>• Les cas font généralement l'objet d'un choix raisonné, ce qui complique la généralisation des conclusions</li> <li>• Moins utile pour les études d'attribution</li> </ul>
<b>Exemples</b>	
L'évaluation de l'aide humanitaire danoise en 2015 a eu recours à deux études de cas approfondies sur le terrain (Soudan du Sud et crise syrienne) ainsi qu'à une revue documentaire plus limitée (Afghanistan) (Mowjee <i>et al.</i> , 2015).	



Etude de processus (non-expérimentale)	
 <b>Définition :</b> Une étude de processus compare la manière dont fonctionnent les processus avec le fonctionnement prévu.	
<b>Utilisation dans l'EAH</b>  Les études de processus sont utilisées pour vérifier si les processus respectent des normes plus larges comme Sphère ou les normes propres à une organisation.	
<b>Conseils</b>  La plupart des guides portant sur les études de processus se concentrent sur les études de processus dans le monde des affaires. Nous n'avons trouvé aucun guide propre à l'action humanitaire.	
Points forts	Points faibles
<ul style="list-style-type: none"> <li>• Pertinente pour répondre à des questions normatives</li> <li>• Pertinente pour répondre à des questions liées à des hypothèses de planification</li> <li>• Pertinente pour répondre à des questions liées à l'efficacité des processus</li> <li>• Peut être menée à tout moment d'une intervention</li> </ul>	<ul style="list-style-type: none"> <li>• Accorde généralement une faible attention à l'impact</li> <li>• Pertinente seulement pour certaines questions</li> <li>• Part du principe qu'un processus correct garantira de bons résultats</li> </ul>
<b>Exemples</b>  L'étude réalisée par le HCR en 2013 sur les diagnostics participatifs est un exemple d'étude de processus. L'étude de processus du Fonds humanitaire commun en Somalie en est un autre (Willitts-King <i>et al.</i> , 2012).	

## Etude de résultats (non-expérimentale)



### Définition :

Une étude de résultats compare les résultats avec les résultats prévus.

### Utilisation dans l'EAH

Utilisation formelle limitée, mais de nombreux rapports d'EAH se concentrent sur les résultats plutôt que sur l'impact.

Points forts	Points faibles
<ul style="list-style-type: none"> <li>• Evite l'exercice souvent difficile qui consiste à évaluer la contribution ou l'attribution des impacts</li> <li>• Se combine bien avec les documents de planification des programmes</li> </ul>	<ul style="list-style-type: none"> <li>• Ne traite pas la question de l'impact. Même si des résultats peuvent être obtenus, ils ne débouchent pas toujours sur des impacts positifs.</li> <li>• Ne précise pas pourquoi les choses se sont produites</li> </ul>

### Exemples

Pas d'exemples formels dans l'EAH, mais l'évaluation Oxfam GB de la réponse à la sécheresse au Kenya (Turnbull, 2012) se réfère à une étude de résultat antérieure des programmes de transfert monétaire.

## Conception participative (non-expérimentale)



### Définition :

L'évaluation participative implique toutes les parties prenantes à toutes les phases de l'évaluation, de la planification initiale à la mise en œuvre des recommandations (voir aussi les éléments sur l'évaluation participative à la Section 4).


### Utilisation dans l'EAH

Aucun véritable exemple n'a été trouvé pour l'EAH. Même si de nombreuses évaluations s'efforcent d'intégrer la participation dans leur travail, nous n'en avons trouvé aucune où l'ensemble du processus a été contrôlé par les parties prenantes. Cette approche est utilisée le plus souvent aux Etats-Unis dans le secteur caritatif. Les méthodes participatives sont fréquemment utilisées dans l'évaluation (Catley, 2009 ; 2014), mais une évaluation n'est considérée comme participative que lorsque les parties prenantes définissent les questions d'évaluation.

### Conseils

USAID a publié une petite note indicative sur l'évaluation participative (USAID, 1996) et le CRS un manuel plus détaillé (Aubel, 1999). Un bon exemple d'évaluation participative est celle de l'approche intégrée Age, Genre et Diversité du HCR en Colombie (Mendoza et Thomas, 2009).

Points forts	Points faibles
<ul style="list-style-type: none"> <li>• Autonomise les participants</li> <li>• Soutient l'utilisation</li> <li>• Promeut le partenariat</li> </ul>	<ul style="list-style-type: none"> <li>• Difficile de prédire les résultats</li> <li>• Sujette aux biais des parties prenantes (pour un maintien du financement, par exemple).</li> </ul>

<b>Série chronologique interrompue (quasi-expérimentale)</b>	
 <b>Définition :</b> La conception en série chronologique interrompue fournit une estimation de l'impact d'une intervention en examinant une série chronologique de données antérieures et postérieures à une intervention.	
<b>Utilisation dans l'EAH</b>	
Peu utilisée actuellement.	
<b>Conseils</b>	
Cochrane fournit un court exemple détaillé de conception en série chronologique interrompue (EPOC, 2013). Glass (1997) propose de nombreux autres exemples de son utilisation. Gilmour <i>et al.</i> (2006) donnent un exemple où la date exacte de l'intervention n'est pas claire. Wagner et Ross-Degnan (2002) donnent des conseils sur l'utilisation de l'analyse de régression sectorielle ou de la série chronologique interrompue.	
<b>Points forts</b>	<b>Points faibles</b>
<ul style="list-style-type: none"> <li>• Peu chère, puisqu'elle utilise en général des données secondaires</li> </ul>	<ul style="list-style-type: none"> <li>• Sujette à la contamination, en particulier en raison d'effets liés au temps comme la maturation (changements comportementaux au fil du temps) et ainsi de suite</li> <li>• Demande un examen attentif pour éliminer les causes alternatives</li> <li>• Demande de fortes compétences en statistiques</li> </ul>
<b>Exemples</b>	
Il semble qu'aucune EAH n'ait utilisé la conception en série chronologique interrompue, mais elle a été utilisée pour des études d'une autre nature dans des contextes humanitaires, notamment une sur le suicide après le tremblement de terre à Taiwan en 1999 (Xang <i>et al.</i> , 2005) ou sur la pneumonie après la catastrophe de Fukushima (Daito <i>et al.</i> , 2013).	

## L'écart dans les différences (quasi-expérimentale)



### Définition :

Cette conception estime l'effet de l'assistance en comparant dans le temps l'évolution de la moyenne du résultat d'intérêt entre le groupe assisté et un groupe de comparaison.

### Utilisation dans l'EAH

Pas d'exemple d'utilisation dans l'EAH. Certaines évaluations se réfèrent à cette méthode mais ne l'appliquent pas de façon rigoureuse.

#### Points forts

- Peut être peu onéreuse si l'on utilise des données secondaires
- Là où des changements importants sont observés, cette conception fournit des illustrations attrayantes de la différence
- Est bien plus intuitive que de nombreuses autres méthodes statistiques

#### Points faibles

- Demande de fortes compétences en statistiques
- Le groupe de comparaison doit être suffisamment similaire

## Comparaison avant/après (quasi-expérimentale ou non-expérimentale)



### Définition :

Une comparaison avant/après compare la situation d'un groupe avant et après l'intervention.

### Utilisation dans l'EAH

Peu utilisée actuellement.

#### Points forts

- Montre qu'un certain changement s'est produit de façon indéniable
- Montre l'ampleur du changement (le niveau d'effet)
- Peut être peu onéreuse si des données secondaires sont disponibles

#### Points faibles

- Pas de groupe de comparaison, ainsi tout changement observé peut être dû à d'autres facteurs, comme l'économie externe, etc.

### Exemples

Durkin *et al.* (1993) ont fait une bonne utilisation des données de référence pour une étude avant/après liée aux survivants de l'inondation au Bangladesh. En utilisant les données issues d'une étude sans rapport réalisée six mois plus tôt, ils ont comparé le comportement des enfants mesuré à ce moment-là avec leur comportement après les inondations, suggérant ainsi que les enfants montraient certains signes de stress post-traumatique.

## Groupe de comparaison (quasi-expérimentale ou non-expérimentale)



### Définition :

Les conceptions de groupe de comparaison comparent le groupe assisté avec un groupe de comparaison sélectionné.

### Utilisation dans l'EAH

On trouve parfois des groupes de comparaison de faible qualité dans l'EAH. Des groupes de comparaison de meilleure qualité peuvent être mis en place.

### Points forts

- Les groupes de comparaison réduisent le risque de prendre les changements de fond pour l'impact de l'intervention
- Les groupes de comparaison peuvent rendre les résultats plus convaincants.

### Points faibles

- Très difficile d'éviter la contamination dans les contextes humanitaires
- De fortes compétences en statistiques et des bonnes données sur le groupe assisté sont nécessaires pour les méthodes comme l'appariement par score de propension

## La conception de discontinuité de la régression (quasi-expérimentale)



### Définition :

Une conception de discontinuité de la régression compare les lignes de régression pour la variable d'intérêt avec la note sur laquelle s'est fondée l'intervention.

### Utilisation dans l'EAH

Nous n'avons trouvé aucun véritable exemple de conceptions de discontinuité de la régression, même si certaines études de discontinuité de traitement se définissaient comme des conceptions de discontinuité de la régression.

### Conseils

Jacob *et al.* (2012) fournissent des indications sur la conception tout comme Better Evaluation et Khander *et al.* (2010: ch.7).

### Points forts

- Convient bien à l'action humanitaire car les groupes assisté et de comparaison sont choisis en fonction des besoins
- Peut produire des illustrations convaincantes
- Plus solide qu'une simple comparaison des valeurs supérieures et inférieures au seuil

### Points faibles

- Demande de fortes compétences en statistiques
- Risque de contamination si les critères d'exclusion ne sont pas appliqués de façon stricte
- Les résultats sont généralisables seulement autour de la valeur du seuil

## Comparaison de discontinuité du traitement (quasi-expérimentale)



### Définition :

Une conception de discontinuité du traitement compare le groupe situé juste sous le point d'exclusion de l'assistance avec le groupe situé juste au-dessous.

### Utilisation dans l'EAH

De temps en temps mais certaines études portent de façon inadéquate le nom de conceptions de discontinuité de la régression.

### Points forts

- Pas de problèmes éthiques
- Convient bien à l'action humanitaire car les groupes assistés et de comparaison sont choisis en fonction des besoins
- Ne demande pas de fortes compétences en statistiques

### Points faibles

- Risque de contamination si les critères d'exclusion ne sont pas appliqués de façon stricte
- Les résultats sont généralisables seulement autour de la valeur du seuil

### Exemples

L'étude de Lehmann et Masterson (2014) sur les programmes monétaires au Liban a utilisé cette conception.

## Conception longitudinale (quasi-expérimentale ou non-expérimentale)



### Définition :

Les études longitudinales effectuent des mesures répétées de la même population sur plusieurs années.

### Utilisation dans l'EAH

Peu utilisée, sauf pour l'analyse des données secondaires de telles études. L'utilisation la plus courante est faite dans les études de cohorte portant sur la santé dans les pays développés.

### Points forts


- Solide indication des tendances progressives au fil du temps

### Points faibles

- Réaliser de telles études coûte cher.
- Identifie les changements et les tendances mais pas les raisons sous-jacentes

### Exemples

Khoo (2007) a utilisé les enquêtes longitudinales des migrants vers l'Australie pour étudier la participation sanitaire et économique des migrants humanitaires en utilisant la régression logistique et deux cohortes comparables.

L'essai contrôlé randomisé (RCT) (conception expérimentale)	
 <b>Définition :</b> Un essai contrôlé randomisé (RCT) compare deux groupes sélectionnés de façon aléatoire, l'un des deux recevant l'assistance mais pas l'autre (le groupe de contrôle).	
<b>Utilisation dans l'EAH</b>	
Très peu d'exemples à cause de problèmes éthiques, du coût et de la complexité. Parfois, les RCT ont été utilisés en phase de relèvement, ou ont comparé différents « packs » d'assistance plutôt que d'utiliser un groupe de contrôle sans assistance.	
Points forts	Points faibles
<ul style="list-style-type: none"> <li>• Certains pensent que c'est la conception la plus rigoureuse</li> <li>• Pertinente pour répondre à des questions causales liées à l'attribution</li> <li>• Grande confiance quant au fait que la réponse reflète de façon précise la vérité du contexte</li> <li>• L'assignation aléatoire résout grandement les problèmes de biais de la sélection</li> </ul>	<ul style="list-style-type: none"> <li>• Nombreux problèmes éthiques (dont l'assignation aléatoire de l'assistance plutôt qu'en fonction des besoins)</li> <li>• Conception la plus chère</li> <li>• Conception la moins flexible</li> <li>• Conditions préalables exigeantes</li> <li>• Adaptée seulement à des types de question spécifiques</li> <li>• Il est rarement possible de maintenir les groupes dans l'ignorance quant au fait qu'ils soient groupe assisté ou groupe de contrôle, ce qui aboutit à de la contamination</li> <li>• Difficile de contrôler la contamination (par exemple, le groupe de contrôle peut recevoir une aide similaire d'une autre source)</li> <li>• Un cadre d'échantillonnage (liste qui comprend plus ou moins l'ensemble de la population) peut être difficile à obtenir dans le contexte humanitaire</li> <li>• Nécessité de choisir de façon aléatoire avant l'intervention les groupes assisté et non-assisté</li> <li>• Caractère généralisable limité car les résultats s'appliquent seulement à cette expérimentation</li> </ul>
<b>Exemples</b>	
Fearon <i>et al.</i> ont réalisé en 2008 une évaluation d'impact de la reconstruction communautaire dans le Comté de Lofa qui utilisait un RCT.	

## Expérimentation naturelle (quasi-expérimentale)



### Définition :

Les conceptions expérimentales naturelles comparent un groupe assisté et un groupe similaire qui, par hasard, n'a pas reçu d'assistance.

### Utilisation dans l'EAH

Parfois utilisée lorsque les chercheurs remarquent que les conditions permettent une expérimentation naturelle.

### Conseils

Les expérimentations naturelles sont décrites dans le chapitre 15 de Remler et Ryzin (2014).<sup>1</sup>

#### Points forts

- Pertinente pour répondre à des questions causales liées à l'attribution
- Grande confiance quant au fait que la réponse reflète de façon précise la vérité du contexte
- Sans problèmes éthiques liés à des expérimentations délibérées
- Peu onéreuse

#### Points faibles

- Problèmes liés au fait que le groupe assisté et le groupe de comparaison peuvent être légèrement différents, et que cela biaise le résultat
- Le chercheur ne peut pas les prévoir de façon anticipée, mais seulement les découvrir si elles se produisent.

### Exemples

Un exemple d'expérimentation naturelle est l'étude portant sur la manière dont les enfants ont été affectés au Nicaragua par l'ouragan Mitch en 1998 (Baez et Santos, 2007). Autres exemples : l'étude de l'impact à long terme de la famine sur les survivants (Meng et Qian, 2009), de l'impact du déstockage sur la nutrition infantile (Boonstra *et al.*, 2001), de l'impact des réfugiés sur l'économie locale (Maystadt, 2011), sur les prises de risques au lendemain de catastrophes (Cameron et Shah, 2012 ; Page *et al.*, 2012) et sur les relations entre les restrictions sécuritaires et la violence (Longo *et al.*, 2014).

Un exemple très simple d'expérimentation naturelle s'est produit après le séisme de Yogyakarta en 2006. Certaines juridictions du gouvernement local étaient déjà en alerte d'une éruption du Mont Merapi et possédaient une infrastructure relative à la préparation aux catastrophes et à leur gestion. Le Fritz Institute a constaté de forts niveaux de satisfaction vis-à-vis de la réponse d'urgence dans ces régions (Bliss et Campbell, 2007).



## 11.4 Les biais

Le choix de la conception, des méthodes et des approches d'échantillonnage influence les types de biais potentiels. Un biais est une menace pour la précision des résultats de l'évaluation (Knox Clarke et Darcy, 2014: 15).

Il existe une différence entre les biais des méthodes qualitative et quantitative. Les biais sont délibérément inclus dans les méthodes quantitatives – les échantillons sont choisis de façon raisonnée, et les chercheurs doivent s'interroger sur l'impact de leurs propres biais et de ceux générés par le processus de recherche. Même dans ce cas, un biais constitue une menace importante pour la validité interne des méthodes quantitatives.

La validité interne désigne dans quelle mesure les conclusions relatives aux causes et aux effets sont justifiées. Un biais peut émerger des préjugés de l'évaluateur et/ou de la conception de l'évaluation. Personne n'est à l'abri d'un biais, mais un biais provenant de la conception de l'évaluation est insidieux car il peut ne pas être manifeste.

### Les biais de conception

Les biais de conception proviennent d'erreurs systématiques de la conception de l'évaluation. Sackett (1979) a identifié 35 différents types de biais susceptibles d'affecter la recherche analytique, y compris diverses formes de biais de sélection et de mesure.<sup>2</sup>

Les biais de sélection se produisent lorsque les éléments d'échantillon retenus sont biaisés d'une manière ou d'une autre. Parmi les exemples, on peut citer celui des listes de distribution aux bénéficiaires utilisées comme cadre d'échantillonnage à partir duquel choisir un échantillon. Cela biaisera l'échantillon si un nombre important de personnes ne recevait pas l'assistance. Utiliser toute autre méthode qu'une véritable sélection aléatoire d'échantillon à partir du cadre d'échantillonnage introduit également le risque de biais.

Les biais de mesure se produisent lorsque la mesure est biaisée d'une manière ou d'une autre. Ils apparaissent pour de multiples raisons. Par exemple, si une balance à fléau n'est pas correctement nivelée, toutes les mesures seront imprécises. Un autre exemple est celui où les personnes interrogées d'une enquête surévaluent le nombre de personnes au sein de leur foyer dans l'espoir d'obtenir plus d'assistance.

Pour les conceptions expérimentales, la sélection aléatoire contribue à éviter les biais de sélection et les dissimulations concernant les personnes qui se trouvent à la fois dans le groupe assisté et le groupe de contrôle aussi bien pour les membres des groupes que pour le personnel qui travaillent avec eux (phénomène dit de « double insu »). Cela aide à éliminer les biais de mesure.<sup>2</sup> Les études quasi-expérimentales sont sujettes à la fois à des biais potentiels de sélection et de mesure, mais ceux-ci peuvent être contrôlés jusqu'à un certain point par des mesures statistiques.

L'échantillonnage raisonné est délibérément biaisé de manière à choisir les cas les plus appropriés pour les questions d'évaluation, mais les conceptions non-expérimentales risquent d'engendrer des biais de sélection non intentionnels en excluant des groupes particuliers (les jeunes, ou les personnes les plus éloignées de la route, par exemple).

Pour les conceptions quasi-expérimentales et non-expérimentales, une approche transparente et rigoureuse de l'échantillonnage, qui s'appuie sur des critères explicites mis en place à l'avance, peut aider à contrôler les biais de conception. Il peut être utile d'inclure un plan d'échantillonnage dans le rapport de démarrage.

### Objet de l'évaluation, source et biais de publication

Les organisations ont davantage tendance à évaluer les programmes considérés comme des succès que comme des échecs. La session de travail organisée en 2009 par l'Association américaine d'évaluation sur les alternatives au format conventionnel contrefactuel a remarqué que « de nombreuses évaluations présentent un biais positif systématique car la plupart des informations sont collectées auprès de personnes qui ont bénéficié du projet » (Bamberger *et al.*, 2009: 2). Le groupe d'évaluation des Nations unies remarque qu'un biais positif provient de ce que les organisations qui mettent en œuvre les activités tirent des bénéfices des projets (UNEG, 2011: 22).

Le rapport L'Etat du système humanitaire (2015: 29) affirme que :

Plus de 50 % des évaluations notent comme « bonne » la performance du sujet de l'évaluation. La plupart des évaluations sont commanditées par des organisations ou des bailleurs ayant un intérêt direct dans les résultats. De ce fait, il est fort possible que le système d'incitations soit orienté vers des conclusions plus positives, même lorsque des consultants externes sont engagés pour réaliser l'évaluation. Une lecture synthétique des résultats nous a menés à la conclusion que « bon » était aussi souvent utilisé pour signifier que la performance était « adéquate ».

Bien sûr, le « cinquante pour cent des évaluations » devrait probablement être qualifié de « cinquante pour cent des évaluations publiées » puisque les évaluations sont sujettes à un biais de publication qui réduit la probabilité selon laquelle des évaluations critiques seront publiées.

Tableau 11.1 : Exemple de discussions en *focus group* et de raisonnement

Groupe	Groupe assisté	Groupe de comparaison	Raisonnement
Homme adulte	3 focus groups	3 focus groups	Une recherche antérieure a montré que les hommes et les femmes accordent une valeur différente à divers éléments de l'assistance.
Femme adulte	3 focus groups	3 focus groups	
Jeune	2 focus groups	2 focus groups	Un groupe pour les garçons et un pour les filles. Certains rapports expriment des inquiétudes en ce qui concerne l'impact de l'assistance sur les jeunes.
Personnes handicapées	1 focus group	1 focus group	Un seul groupe en raison du petit nombre de personnes.
Personnes âgées	1 focus group	1 focus group	Un seul groupe en raison du petit nombre de personnes.

### Les biais relatifs à la collecte de données

L'évaluation de la réponse aux cyclones tropicaux de 2004 dans les Caraïbes (Gamarra *et al.*, 2005: 3) comprenait une petite section à propos de la manière dont l'équipe s'attaquait aux biais, notamment les biais des organisations, les biais géographiques, et les biais de mémoire.

Les dispositifs permettant de contrôler les biais comprennent la triangulation et d'autres approches dont :

- Interroger des personnes situées loin de la route ou qui ont souvent été interrogées. Interroger le plus grand nombre possible de personnes ne faisant pas partie d'organisations, et vérifier les biais de mémoire en les confrontant avec des rapports de situation (Gubbbs et Bouquest, 2013: 18).
- Interroger la population affectée sans que le personnel de l'organisation soit présent. Éviter de poser des questions portant directement sur l'efficacité de l'organisation, utiliser des éléments de preuve indirects sur la relation, la compréhension du projet et les perceptions d'appropriation (Cosgrave *et al.*, 2010: 14).
- Assurer un équilibre hommes-femmes au sein des personnes interrogées de manière à limiter les biais de genre (Barakat *et al.*, 2006: 139).
- Recouper régulièrement les informations auprès d'une diversité de personnes aussi large que possible (Humanitarian Initiatives *et al.*, 2001: Vol. 3, sect. 6,1).

- Choisir un consultant externe comme chef d'une équipe constituée de personnes issues des organisations évaluées (Reed *et al.*, 2005: 1-2). Éviter que toute personne ayant déjà travaillé avec une organisation particulière ne mène les entretiens avec cette organisation (Cosgrave *et al.*, 2008: 21).
- Utiliser des outils collaboratifs pour partager les résultats émergents et les examiner au sein de l'équipe. Identifier ouvertement tout biais potentiel à l'intérieur de l'équipe (Ternström *et al.*, 2007: 90-91).
- Analyser les résultats pays par pays pour éviter tout biais introduit par la surreprésentation de certains pays dans une enquête en ligne (Steets *et al.*, 2013: 15).
- Ne pas informer les enquêteurs de l'objectif du statut du traitement d'une communauté, et ne pas informer la population affectée de l'objectif de l'enquête (Beath *et al.*, 2013: 46).

### Les biais relatifs au genre et à l'âge

Des biais de genre et d'âge peuvent survenir lors de la collecte des données et donner une image trompeuse de ce qui est en train de se produire.

Voir également le document d'orientation de l'UNEG « Intégrer les droits de l'Homme et l'égalité des sexes aux évaluations » (2014).

### Prévenir les biais de genre et d'âge lors de la collecte de données sur le terrain

Les mesures suivantes contribueront à garantir qu'une évaluation est sensible au genre et à l'âge et produit des données pouvant être désagrégées par sexe et par âge :

- Garantir que l'équipe de terrain comprend un nombre égal de femmes et d'hommes. Ce sont généralement les intervieweurs de sexe féminin et les traductrices qui ont le meilleur accès aux femmes et filles de la population affectée.
- Garantir que les informateurs clés comprennent des femmes qui connaissent bien la communauté et les besoins spécifiques des femmes ainsi que la manière dont ils ont été satisfaits si c'est le cas. Cela peut comprendre des enseignantes, des infirmières et des vendeuses influentes sur les marchés. Garantir que certains informateurs clés connaissent bien les besoins des enfants, des jeunes et des personnes âgées.
- Garantir que les femmes composent au moins la moitié des participants des *focus groups* et des entretiens de groupe. Idéalement, les femmes doivent être interrogées à l'écart des hommes qui ont tendance à dominer les discussions où hommes et femmes sont présents.





- Organiser des *focus groups* et des entretiens de groupe en fonction des âges, par exemple avec des enfants, des jeunes et des personnes âgées. Les adultes ont tendance à dominer les discussions où tous les âges sont représentés.
- Tenir un registre du sexe des informateurs clés et des autres personnes interrogées de manière à vérifier tout biais potentiel de genre au niveau des informations tirées des entretiens.

Source : Tiré en partie de Mazurana *et al.* (2011)

## Les biais propres aux évaluateurs

Les biais propres aux évaluateurs peuvent comprendre une aversion à l'égard d'une organisation en particulier ou de l'approche d'un programme, ou la tentation de réprimer des conclusions très critiques par peur de perdre de futurs contrats avec le client. De tels biais peuvent s'appliquer à toutes les formes de conception de recherche, mais les méthodes non-expérimentales présentent le plus grand risque. De tels biais peuvent être contrôlés dans une certaine mesure en utilisant la triangulation entre chercheurs et en comparant les conclusions de différents chercheurs.

Parmi les approches susceptibles d'aider à contrôler les biais propres aux évaluateurs, on peut noter :

- Recruter des évaluateurs qui ont conscience de leurs propres biais et qui essaient de les prendre en compte
- Transparence en matière de prise de décision concernant l'échantillonnage
- Transparence en matière de collecte des données et de stratégies d'analyse
- Liens clairs entre éléments de preuve, résultats et conclusions
- Assurance qualité externe
- Triangulation des résultats entre évaluateurs
- Faire en sorte que les mêmes données soient codifiées par au moins deux évaluateurs et comparer les résultats afin de détecter d'éventuels biais systématiques.



### Astuce :

Soyez transparent en ce qui concerne les biais. Insérez dans le rapport d'évaluation une remarque à propos des potentiels biais de l'équipe d'évaluation et identifiez quelles mesures vous avez prises pour les contrôler ainsi que toute autre forme de biais.

# 12 / L'échantillonnage

Cette section couvre la question de l'échantillonnage. Il est parfois possible d'enquêter sur chaque cas, mais souvent les cas sont trop nombreux ou alors le nombre d'informateurs potentiels est trop important pour que l'on puisse tous les consulter. Au lieu de cela, on utilise un échantillon.

**Définition : Echantillonnage**

Sélection d'un sous-ensemble de population à inclure dans une étude (plutôt que la population dans son ensemble).

Presque toutes les évaluations ont recours à l'échantillonnage. Cette section présente comment même la plus petite évaluation peut utiliser de bons choix d'échantillonnage pour améliorer la qualité de l'évaluation.

Sous sa forme la plus simple, un échantillon est établi par simple convenance à partir de la population affectée, des informateurs clés ou des autorités. Sous leur forme la plus complexe, les échantillons peuvent être sélectionnés de manière aléatoire parmi les différentes couches sociales. Il est attendu qu'une petite évaluation de bonne qualité utilise :

- Un échantillonnage non aléatoire pour la collecte des données qualitatives
- Un échantillonnage aléatoire pour la collecte des données quantitatives

**Définition : Echantillonnage non aléatoire**

L'échantillonnage non aléatoire choisit l'échantillon d'après certaines propriétés de l'échantillon.

Dans les échantillons non aléatoires, la probabilité qu'un membre donné d'une population soit sélectionné dépend de ses caractéristiques plutôt que d'une chance égale. C'est pour cette raison que les échantillons non aléatoires ne sont pas représentatifs de la population dans son ensemble.



### **Définition : Échantillonnage aléatoire**

L'échantillonnage aléatoire établit un échantillon d'une population où chaque membre de la population a une chance égale d'être sélectionné.

Comme chaque membre de la population a une chance égale d'être sélectionné dans un échantillon aléatoire, de tels échantillons sont dits représentatifs.

Cette section examine les facteurs que les évaluateurs doivent prendre en compte en concevant leur stratégie d'échantillonnage. Les évaluateurs doivent avoir une stratégie d'échantillonnage explicite, conçue de façon crédible pour répondre aux questions d'évaluation.

## 12.1 Les problèmes d'échantillonnage courants dans l'EAH

Les problèmes d'échantillonnage suivants sont souvent observés dans l'EAH :

- Trop grand recours aux échantillonnages de disponibilité ou « de convenance » pour les études qualitatives. Cette approche est insuffisante. Morra et Rist (2009: 363) notent que « les échantillons de convenance sont très peu efficaces pour déduire tout type d'élément de preuve ou tendance à partir des données ». L'échantillonnage de convenance de la population affectée est très trompeur si l'assistance a été fournie sur une base de convenance similaire car l'échantillon indiquera que tout le monde a bénéficié de l'assistance.
- Une taille d'échantillon aléatoire insuffisante pour permettre des généralisations statistiques valables à partir de l'échantillon.
- Une utilisation inappropriée de l'échantillonnage aléatoire dans des études qualitatives : Collier et Mahoney (1996 : 57) remarquent que pour les échantillons qualitatifs, « la stratégie visant à éviter les biais de sélection au moyen d'un échantillonnage aléatoire peut créer autant de problèmes qu'elle n'en résout ».
- L'absence de clarification de l'approche d'échantillonnage dans les rapports d'évaluation, en particulier lorsque des méthodes qualitatives sont utilisées.

Daniel (2013) propose une introduction détaillée aux échantillons aléatoires et non aléatoires. Enfin, le *Practical Guide to Sampling* – publié par le National Audit Office de Grande-Bretagne (2000) - constitue une bonne ressource pour se lancer dans l'échantillonnage.

## 12.2 L'échantillonnage non aléatoire

L'échantillonnage aléatoire est souvent inapproprié pour les approches qualitatives qui prédominent dans de nombreuses EAH. Le guide de Daniel sur l'échantillonnage (2013, chapitre 3) avance que l'échantillonnage non aléatoire peut être approprié si :

- La recherche est exploratoire
- Une décision rapide doit être prise
- Des individus spécifiques doivent être ciblés
- L'on souhaite fournir des exemples illustratifs
- L'accès est difficile ou que la population est très dispersée
- Le temps et l'argent sont limités mais que du personnel qualifié et hautement formé est disponible
- Un cadre d'échantillonnage n'est pas disponible
- Des méthodes qualitatives sont utilisées
- L'on doit utiliser des procédures opérationnelles faciles
- La cible est de petite taille.

Les situations d'accès difficile sont courantes dans l'EAH. Lors d'une évaluation réalisée en RDC, les évaluateurs ont constaté que :

A cause du contexte changeant et de la logistique à l'Est de la RDC, il était impossible de sélectionner les sites en avance de façon aléatoire, en particulier dans le Nord-Kivu et le Sud-Kivu, qui ont connu des conflits et des déplacements répétés durant une grande partie de l'année 2012. Dans le même temps, des sites ont été sélectionnés de manière opportuniste dans un délai très court de manière à ce que les membres de l'équipe puissent observer les activités en cours. (Backer *et al.*, 2013: 12)

Dans le cadre de recherches qualitatives, il est rare que l'on ait besoin d'échantillons représentatifs ou d'interroger un responsable ayant travaillé pendant la durée de présence moyenne, mais il est utile d'interroger le responsable qui est resté le plus longtemps et qui peut fournir le plus grand nombre d'informations. Si l'on veut savoir quels facteurs ont contribué au succès d'un programme de microcrédit, et que l'on utilise une approche qualitative, l'échantillon est biaisé si l'on interroge les utilisateurs pour qui cela a le mieux fonctionné et qu'on leur demande ce qui a permis leur succès. On peut également interroger les utilisateurs pour qui cela a le moins bien fonctionné afin d'observer ce qui a mené à cet échec. Par conséquent, il est impossible de généraliser à l'ensemble de la population mais on peut avancer quelques éléments quant aux facteurs de succès.

L'échantillonnage raisonné est probablement le type d'échantillonnage le plus sérieux pour des études qualitatives parce que les membres de l'échantillon sont choisis de façon délibérée pour les connaissances qu'ils peuvent apporter à la recherche. C'est pourquoi la saturation des données est utilisée comme le critère permettant de définir la taille de l'échantillon dans les échantillons non aléatoires.



**Définition : Saturation des données**

La saturation des données se produit lorsque de nouveaux cas n'apportent plus de nouvelles connaissances.

Le risque propre aux échantillons qualitatifs n'est pas la non représentativité, mais qu'ils n'incluent pas toutes les principales catégories.

## Les types de stratégie d'échantillonnage non aléatoire

Daniel (2013) classe les types d'échantillonnage non aléatoire en quatre catégories : de disponibilité, raisonné, par quotas et facilité par les personnes interrogées. La stratégie d'échantillonnage appropriée dépend de la méthode et du type de question à laquelle l'équipe essaie de répondre.

### L'échantillonnage de disponibilité

**Définition : L'échantillonnage de disponibilité**

L'échantillonnage de disponibilité est une procédure d'échantillonnage où la sélection est fondée sur la disponibilité des échantillons, la convenance du chercheur ou l'auto-sélection.

On l'appelle communément échantillonnage de convenance et Daniel réserve ce nom à un sous-type. Alors que Patton affirme que « l'échantillonnage de convenance n'est ni stratégique ni raisonné ; il est le fruit de la paresse et en grande partie inutile » (2014: Module 39), Daniel remarque qu'il est « la procédure d'échantillonnage la plus fréquemment utilisée dans le cadre d'une recherche ». L'échantillonnage de disponibilité présente l'avantage d'être peu onéreux, simple et pratique mais il a pour principales faiblesses d'être la méthode non aléatoire la moins fiable, de sureprésenter les échantillons les plus disponibles et de tendre à une sous-estimation de la variabilité au sein de la population. Ce type d'échantillonnage doit être utilisé seulement lorsqu'aucune alternative n'est envisageable.

La pire combinaison possible revient à utiliser l'échantillonnage de convenance lorsque l'assistance a également été fournie sur la base de la convenance.



## L'échantillonnage raisonné

### Définition : L'échantillonnage raisonné

L'échantillonnage raisonné sélectionne l'échantillon sur une base intentionnelle de façon à ce que les éléments échantillonnés puissent fournir le plus d'informations possibles à l'étude.

Les échantillons raisonnés peuvent être sélectionnés en se fondant sur :

- La réputation et l'expérience. L'équipe peut par conséquent préférer parler avec les responsables humanitaires les plus expérimentés ou à la population affectée la plus exposée aux activités en train d'être évaluées.
- Le fait qu'ils sont jugés comme les mieux placés pour fournir les informations permettant de répondre aux questions d'évaluation. Par exemple, si une question d'évaluation cherche à savoir si les groupes vulnérables ont été traités avec dignité, les membres des groupes vulnérables sont les mieux placés pour répondre.
- Le fait qu'ils sont jugés comme probablement les mieux à même de démontrer une composante particulière de la théorie du changement (TdC). De tels cas peuvent infirmer des composantes de la TdC présumée.
- Le fait qu'ils sont considérés comme ayant vécu l'expérience la plus représentative. L'échantillonnage raisonné est utilisé pour les questions descriptives qui cherchent à savoir quelle a été l'expérience la plus courante.
- Le fait qu'ils sont considérés comme des exceptions par rapport à l'expérience la plus représentative. Ces cas sont souvent extrêmement importants pour les questions de causalité, parce que les aberrations peuvent être des cas infirmant la TdC sous-jacente. Par exemple, une évaluation d'un programme alimentaire peut prendre un échantillon de mères qui ont abandonné le programme tôt ainsi que de mères ayant poursuivi le programme le plus longtemps possible.
- Leur diversité, de façon à ce que le groupe soit inclusif. Ainsi, il faudrait essayer d'inclure des familles monoparentales et des familles nucléaires, des jeunes et des personnes âgées, et ainsi de suite. Pour des recherches qualitatives, les échantillons sont souvent choisis pour être inclusifs plutôt que représentatifs. Cela s'explique par le fait que chaque cas doit apporter du sens et des données à la recherche.
- Leur capacité à indiquer si la théorie du changement est correcte ou non. Ces types de cas sont à nouveau extrêmement importants pour confirmer ou infirmer des théories particulières de causalité. Par exemple, si la TdC énonçait que l'amélioration de l'état nutritionnel dépendait non seulement de l'alimentation mais aussi de l'hygiène et de l'accès aux soins de santé, il est possible d'établir des échantillons qui respectent seulement certaines de ces conditions pour tester la TdC.

## L'échantillonnage par quotas



### Définition : L'échantillonnage par quotas

L'échantillonnage par quotas divise la population en sous-catégories qui s'excluent mutuellement, et collecte ensuite les informations selon une taille d'échantillon précédemment établie ou une proportion pour chaque catégorie.

Par exemple, si l'on échantillonne des ménages au sein d'une population qui a reçu de l'assistance, on peut définir un quota en utilisant au moins deux des éléments suivants :

- Ménages avec des jeunes enfants
- Ménages avec uniquement des parents âgés
- Ménages avec une femme chef de famille
- Ménages avec un homme chef de famille
- Ménages avec un enfant chef de famille
- Ménages aisés
- Ménages pauvres
- Ménages habitant près de la route
- Ménages habitant loin de la route.
- Les échantillons établis au sein de chaque quota peuvent être élaborés en fonction de la disponibilité, d'un choix raisonné ou même avec l'aide des personnes interrogées. L'échantillonnage par quotas peut contribuer à garantir que les échantillons comprennent suffisamment de femmes, d'enfants, de personnes âgées, ou toute autre catégorie d'intérêt.



### Astuce :

Utilisez des quotas si vous devez utiliser un échantillonnage de disponibilité

Si vous n'avez pas d'autre option que d'utiliser un échantillonnage de disponibilité, introduisez des quotas de façon à ce que l'échantillonnage soit au moins plus inclusif. Vous pouvez utiliser des quotas dans l'échantillonnage de disponibilité en faisant en sorte qu'une personne interrogée sur deux soit une femme, qu'une personne interrogée sur cinq soit un jeune homme, etc.

### L'échantillonnage facilité par les personnes interrogées

L'utilisation la plus courante d'échantillonnage facilité par les personnes interrogées dans l'EAH est l'utilisation d'échantillonnage par effet « boule de neige » dans le cadre d'entretiens avec des informateurs clés. Dans un échantillonnage en boule de neige, chaque personne interrogée doit suggérer des noms d'autres personnes (Goodman, 1961) susceptibles d'avoir un avis intéressant sur le sujet. Et ainsi de suite jusqu'à ce qu'aucun nouveau nom n'émerge. Voir Biernacki et Waldorf (1981) pour des éléments utiles sur cette technique.

Schéma 12.1 : L'échantillonnage par effet « boule de neige »



Les renvois à des informateurs-clés déjà interrogés et les doublons ne sont pas indiqués.

L'échantillonnage par effet « boule de neige » fait l'objet d'un certain nombre de critiques (Erickson, 1979) et d'autres approches sont parfois préconisées, notamment :

- L'échantillonnage mené par les personnes interrogées (où la personne qui suggère des contacts et ces contacts reçoivent une rémunération) (Heckathron, 1997 ; Wang, 2005)
- L'échantillonnage ciblé, qui s'appuie sur une recherche initiale utilisant une cartographie ethnographique pour identifier la population échantillon (Watters et Biernacki, 1989).

De telles critiques à l'encontre de l'échantillonnage par effet « boule de neige » et des alternatives s'appliquent davantage à l'échantillonnage des populations cachées comme les usagers de drogues par voie intraveineuse qu'à celui des gestionnaires de l'aide. L'échantillonnage par effet « boule de neige » peut souvent mener à un très grand nombre d'entretiens. C'est enfin une bonne technique pour surmonter tout biais non intentionnel de sélection au niveau des cibles initiales d'entretien.

La recherche menée par Guest *et al.* (2006) sur le nombre d'entretiens qualitatifs nécessaires pour arriver à saturation formule les constats suivants :

1. L'utilisation de l'échantillonnage par effet « boule de neige » pour identifier les personnes à interroger peut s'avérer inefficace car peu de nouveaux thèmes émergent après 6 à 12 entretiens avec un groupe spécifique.
2. La traditionnelle question de fin d'entretien – « A qui d'autre pensez-vous que nous pourrions parler de ce sujet ? » – doit probablement être complétée par une autre : « Nous interrogeons les groupes A, B, et C. Quels autres groupes pensez-vous qu'il serait utile d'interroger ? »



**Astuce :**

Classez les personnes interrogées en groupes et cherchez à avoir des groupes de 6 à 12 personnes tout en continuant à vérifier si de nouveaux sujets et de nouvelles données continuent à émerger.

## La taille des échantillons non aléatoires

Le chercheur qui travaille sur des échantillons qualitatifs doit néanmoins prendre de nombreuses décisions relatives à l'échantillonnage, non seulement en ce qui concerne la stratégie d'échantillonnage mais aussi la taille des échantillons. De manière générale, l'indication du nombre adéquat d'échantillons qualitatifs est celle de la *saturation des données*. Toutefois, il existe peu de documents d'orientation opérationnelle sur le nombre d'entretiens nécessaires pour atteindre cette saturation.

L'utilisation d'un outil de collecte des éléments de preuve comme le tableau de synthèse des observations décrit dans la [Section 8 : La phase de démarrage](#), où les éléments de preuve sont enregistrés selon un système de codage préétabli, peut aider à identifier la saturation. Dans ce but, un tel outil de collecte peut être amélioré en enregistrant si une remarque faite lors d'un entretien particulier ou par une source particulière est nouvelle ou renforce simplement des remarques déjà notées. Cela aidera à identifier plus rapidement la saturation des données.

Creswell et Clark (2011 : 174) suggèrent que quatre à six cas sont adéquats pour une étude de cas et vingt à trente entretiens pour une étude reposant sur des entretiens. Ces recommandations semblent toutefois des règles empiriques fondées sur la pratique des auteurs.

Pour des entretiens qualitatifs de ménages, Few *et al.* (2014: 58) suggèrent un minimum de vingt entretiens pour une communauté de 500 ménages et un minimum de cinquante pour une communauté de 5.000 ménages de manière à garantir qu'une gamme adéquate d'opinions est prise en compte. Few *et al.* suggèrent également de stratifier cet échantillon pour garantir la couverture de différents groupes sociaux et que les ménages soient choisis au hasard pour réduire les biais. Il faut enfin noter que les échantillons de cette taille sont rarement généralisables à l'ensemble de la population.

Pour les *focus groups*, Morgan (1997) remarque que peu de thèmes nouveaux émergent après le troisième groupe. Une étude portant sur la saturation des données lors d'entretiens d'informateurs clés (Guest *et al.*, 2006) a conclu que la saturation des données avait été atteinte entre le sixième entretien et le douzième. Dans les deux cas, Morgan et Guest se réfèrent à des groupes homogènes. La plupart des interventions humanitaires impliquent diverses parties prenantes, et la voix de chacune doit être entendue. Par conséquent, si vous observez l'éducation et les jeunes, il se peut que vous ayez trois *focus groups* comprenant chacun des jeunes hommes, des jeunes femmes, des enseignants et des parents. De la même manière, il se peut que vous ayez besoin de réaliser une série d'entretiens avec des chefs communautaires locaux, des représentants du groupe des femmes, du personnel terrain des ONG, des responsables d'ONG, etc.

En matière d'évaluation, la crédibilité est une question clé. Une évaluation ne peut être utile que si elle est jugée crédible. Même si des entretiens d'informateurs clés réalisés avec un échantillon de douze personnes issues d'un groupe donné peuvent arriver à saturation des données, ce n'est pas le seul facteur à prendre en compte.

Des entretiens supplémentaires peuvent :

- Rendre l'évaluation plus crédible en ayant cherché à connaître l'opinion d'un plus grand nombre de personnes.
- Renforcer les sujets provenant des douze premiers entretiens, en apportant une triangulation.

La triangulation est abordée dans la Section 13 : Les méthodes de terrain. La triangulation est un élément critique des méthodes qualitatives et combinées, ce qui signifie que les évaluations doivent aller au-delà du nombre d'entretiens et de cas nécessaires pour arriver à une simple saturation des données.

Une approche peut consister à utiliser des entretiens de groupe, des ateliers ou des revues après action comme manières plus efficaces de consulter un plus grand nombre de personnes, en plus des entretiens avec les informateurs clés. Dans ce cas, les évaluateurs doivent s'efforcer d'arriver à saturation des données pour chacune des méthodes utilisées. Si des méthodes différentes identifient les mêmes sujets et problèmes, cela constitue une forme de triangulation des résultats de l'évaluation.

### Qu'est-ce qui ne va pas dans le fait d'utiliser un échantillonnage aléatoire pour des méthodes qualitatives ?

A première vue, l'utilisation d'un échantillonnage aléatoire devrait aider à prévenir les biais en échantillonnant à la fois les études qualitatives et les études quantitatives. Toutefois, selon Gerring (2007: 87-88), alors qu'une collection de petits échantillons aura généralement la même valeur moyenne que la population, les échantillons individuels peuvent avoir des valeurs moyennes très différentes de celles de la population. Par conséquent, les échantillons qualitatifs aléatoires peuvent fournir une estimation particulièrement trompeuse de la situation dans laquelle se trouve l'ensemble de la population.

Avec les méthodes quantitatives, le fait que chaque nouveau cas ajoute seulement une faible quantité d'informations n'est pas un problème parce que l'échantillon comprend de nombreux cas. C'est un problème en revanche pour les études qualitatives où l'on attend que chaque cas contribue de façon significative à la connaissance acquise.

Le dernier problème est le suivant : parce que l'échantillon a été sélectionné de façon aléatoire, on peut partir du principe erroné que les résultats peuvent être généralisés. Or, ceci n'est pas possible si la taille de l'échantillon est trop petite.



#### Astuce :

Si vous devez sélectionner seulement quelques cas, n'utilisez pas l'échantillonnage aléatoire ou pseudo-aléatoire mais choisissez les cas de façon raisonnée ou en utilisant des quotas.

## 12.3 L'échantillonnage aléatoire

L'échantillonnage aléatoire est essentiel pour les méthodes quantitatives où il est nécessaire de généraliser l'échantillon à l'ensemble de la population. Si des échantillons sélectionnés ne sont pas utilisés, il est impossible de faire des généralisations valides à partir de l'échantillon pour l'ensemble de la population car l'échantillon n'est pas représentatif. Un véritable échantillonnage aléatoire requiert un cadre d'échantillonnage et une liste de l'ensemble de la population à partir de laquelle l'échantillon aléatoire peut être établi. Il existe rarement une liste complète, mais on peut en trouver certaines approximatives, comme les listes de bénéficiaires, à partir desquelles l'échantillon peut être établi.

### L'échantillonnage pseudo-aléatoire

Dans de nombreux cas, des méthodes pseudo-aléatoires peuvent être utilisées à la place d'un véritable échantillonnage aléatoire. C'est même courant lorsqu'il n'existe aucun cadre d'échantillonnage complet. Les méthodes pseudo-aléatoires impliquent de prendre un point de départ aléatoire (ou pseudo-aléatoire) ??? et de réaliser l'échantillonnage en utilisant un intervalle à partir de ce point.



#### **Définition : L'échantillonnage pseudo-aléatoire**

Echantillonnage retenu lorsqu'il n'existe aucun cadre d'échantillonnage. De manière générale, le premier choix est sélectionné de façon aléatoire à partir d'un point de départ sélectionné de façon raisonnée, et les choix suivants sont sélectionnés en utilisant la même règle.

Diverses règles peuvent s'appliquer, comme choisir le point situé au centre d'un groupe sur une carte ou une photographie aérienne avant d'avancer dans la direction définie par un stylo que l'on a fait tourner sur le sol comme une toupie. Deux règles d'intervalles s'appliquent : soit les règles d'intervalles fixes (par exemple, sélectionner une maison toutes les dix maisons – en éliminant neuf maisons), soit les règles d'intervalles variables qui utilisent une liste de nombres aléatoires ou un générateur de nombres aléatoires.



#### **Astuce :**

Utiliser une application de nombres aléatoires sur votre smartphone pour générer des valeurs d'intervalles aléatoires, par exemple, si vous voulez interroger un foyer sur dix, vous définirez en moyenne les valeurs de nombres aléatoires entre 1 et 17 pour générer des intervalles aléatoires avec une moyenne de 9.

Il est important d'être conscient que l'inférence statistique est fondée sur un véritable échantillonnage aléatoire et que l'utilisation de méthodes pseudo-aléatoires peut introduire un biais inconnu.



## L'échantillonnage par grappes et l'effet de conception



### Définition : L'échantillonnage par grappes

L'échantillonnage par grappes est un échantillonnage où un certain nombre de sites sont échantillonnés, chacun comportant un nombre spécifique de cas.

Le principal avantage de l'échantillonnage par grappes est qu'il peut être utilisé sans cadre d'échantillonnage des individus au sein de la population.

Une technique courante d'échantillonnage par grappes consiste à choisir 30 grappes de 30 individus mais d'autres techniques sont possibles. L'échantillonnage par grappes requiert un plus grand échantillon global mais c'est une technique efficace en l'absence d'un bon cadre d'échantillonnage ou lorsque la population est dispersée sur un vaste zone. Dans ce dernier cas, l'échantillonnage par grappes permet un travail de terrain plus rapide et plus efficient. L'échantillonnage par grappes est largement utilisé pour la nutrition et les enquêtes de vaccination. L'échantillonnage par grappes peut utiliser un échantillonnage aléatoire ou raisonné pour sélectionner les grappes, et un échantillonnage aléatoire ou pseudo-aléatoire au sein de la grappe.

Utilisez un calculateur d'échantillons pour réaliser un calcul précis de la taille de l'échantillon que vous souhaitez, et ajoutez ensuite une marge de sécurité d'au moins 10 % de non réponses. Si des enquêtes répétées sont prévues, vous devez prendre en compte les taux de défection entre enquêtes.



### Définition : Effet de conception

L'effet de conception (également appelé « effet de plan ou effet du plan de sondage ») est le facteur par lequel vous devez modifier votre taille d'échantillon lorsque vous vous écarterez d'un simple échantillonnage aléatoire.

Un effet de conception de deux (doublant la taille de l'échantillon) est en général utilisé pour les enquêtes nutritionnelles utilisant un échantillonnage par grappes (IFAD, 2014), mais une étude portant sur les effets de conception des échantillons par grappes (Kaiser *et al.*, 2006) a révélé qu'un effet de conception de 1,5 était adéquat pour la plupart des enquêtes de malnutrition aiguë. La même étude a montré que les enquêtes de mortalité due à un conflit utilisant un échantillonnage par grappes avaient des effets de conception de quatre ou plus. Pour leur étude de mortalité durant la crise alimentaire de 2000 en Ethiopie, Salama *et al.* (2001: 564) ont utilisé un effet de conception de quatre pour la mortalité.

L'effet de conception provient de ce que chaque nouveau membre d'une grappe fournit moins d'informations nouvelles qu'un nouveau membre indépendant et cela, en raison des similitudes entre membres de chaque grappe. Ainsi, l'effet de conception est plus élevé pour les situations dans lesquelles les membres des grappes sont plus susceptibles d'être semblables.

## L'échantillonnage stratifié

La seule stratégie d'échantillonnage présentant un effet de conception inférieur à 1 est l'échantillonnage aléatoire stratifié pour lequel l'échantillon peut être légèrement plus petit que pour l'échantillonnage aléatoire standard.



### Définition : L'échantillonnage stratifié

L'échantillonnage stratifié est une approche d'échantillonnage où la population est divisée en segments s'excluant mutuellement avant qu'un simple échantillon aléatoire ne soit sélectionné au sein de chacun d'eux.

L'échantillonnage stratifié peut être utile si l'on craint que des minorités spécifiques ne soient oubliées.

## Calculer la taille nécessaire des échantillons aléatoires

Calculer la taille nécessaire de l'échantillon dépend de si l'on veut généraliser un échantillon relatif à une population ou comparer deux populations.

Il dépend également de si vous observez les proportions au sein d'une catégorie particulière (malnutris, non malnutris ; ayant reçu un kit abri, n'ayant pas reçu un kit abri) ou les valeurs moyennes de certaines variables (revenus des foyers, pourcentage de la nourriture d'urgence revenue, litres d'eau utilisés par jour).

Il est faux de croire que la taille de l'échantillon doit être proportionnelle à la population, même si des échantillons plus petits peuvent être utilisés avec des populations de petite taille. Daniel suggère qu'il est pertinent de corriger la taille réelle de l'échantillon quand celui-ci représente plus de 5 % de la population.

Une fois que la population descend en dessous de 400 personnes environ, la taille de l'échantillon nécessaire dépasse les 50 % de la population (dans le pire des cas). Faire un recensement devient plus intéressant lorsque la taille de l'échantillon augmente en proportion d'une population.

**Astuce :**

Utilisez un calculateur en ligne de taille d'échantillon pour calculer la taille de l'échantillon dont vous avez besoin. Les calculateurs en ligne ci-dessous présentent l'avantage de donner pour chaque formule les définitions des termes utilisés.

- Calculer une statistique sommaire pour une population : [www.select-statistics.co.uk/sample-size-calculator-proportion](http://www.select-statistics.co.uk/sample-size-calculator-proportion)
- Calculer la moyenne d'une population : [www.select-statistics.co.uk/sample-size-calculator-mean](http://www.select-statistics.co.uk/sample-size-calculator-mean)
- Comparer les proportions au sein de deux groupes : [www.select-statistics.co.uk/sample-size-calculator-proportion](http://www.select-statistics.co.uk/sample-size-calculator-proportion)
- Comparer les moyennes au sein de deux groupes : [www.select-statistics.co.uk/sample-size-calculator-two-means](http://www.select-statistics.co.uk/sample-size-calculator-two-means)

Lorsque nous calculons la valeur de certains paramètres d'une population à partir d'un échantillon, nous réalisons un calcul qui a une certaine probabilité d'échouer avec un intervalle de confiance donné.

**Définition : L'intervalle de confiance**

L'intervalle de confiance est la fourchette à l'intérieur de laquelle on s'attend à trouver la valeur de la population dans son ensemble.

L'intervalle de confiance est généralement exprimé sous la forme d'un pourcentage (+/-) : +/- 5 % est le plus courant. Il est également appelé marge d'erreur. Nous ne pouvons jamais être totalement sûrs que la valeur d'une population tombera dans cet intervalle de confiance. Dans un petit pourcentage de cas, il peut tomber en dehors.

**Définition : Le niveau de confiance**

Le niveau de confiance est la probabilité que la valeur de la population dans son ensemble ne tombe dans l'intervalle de confiance.

Un niveau de confiance de 95 % signifie qu'il y a 5 % de chance que la valeur de la population tombe à côté de l'intervalle de confiance – c'est-à-dire 5 % de chance d'obtenir un résultat positif faux.

Lorsque nous calculons une moyenne au sein d'une population, il faut prendre en compte la variabilité de la population (calculée à partir de la variabilité de l'échantillon).

**Définition : L'écart standard**

L'écart standard est une mesure de la variabilité d'un paramètre.<sup>3</sup>

Deux autres facteurs sont importants pour comparer des groupes : la taille d'effet et la puissance statistique.

**Définition : La taille d'effet**

La taille d'effet est la différence proportionnée entre la variable étudiée du groupe traité et celle du groupe de contrôle.

Un problème courant de l'EHA est que les concepteurs surestiment souvent l'effet de leur intervention. De très grands échantillons sont nécessaires lorsque le niveau d'effet est petit. À l'inverse, un grand niveau d'effet demande seulement des petits échantillons.

**Définition : La puissance statistique**

La puissance statistique est la probabilité qu'un résultat négatif ne soit pas un résultat négatif faux (1 moins le risque d'un résultat négatif faux).

Une puissance de 80 % signifie que la probabilité de ne pas détecter une différence significative est de 20 % (en d'autres termes - s'il y a une différence significative, c'est un résultat négatif faux). Lors d'essais médicaux, les faux résultats négatifs sont généralement plus acceptables que les faux résultats positifs, d'où la marge d'erreur généralement plus importante. On peut toutefois se demander si les évaluations réalisées pour déterminer si les interventions ont eu un impact quelconque peuvent utiliser une puissance aussi faible. Une puissance de 90 % (avec 10 % de chance d'un résultat négatif faux) peut être plus appropriée. Le guide d'échantillonnage du Food and Nutrition Assistance Project suggère d'utiliser une puissance de 90 % à chaque fois que les ressources le permettent (Magnani, 1997).

Tableau 12.1 : Facteurs influençant la taille de l'échantillon

	Calculer une statistique sommaire pour une population		Comparer deux populations (entre deux groupes ou avant/après)	
	Calculer une proportion	Calculer une moyenne	Comparer deux proportions	Comparer deux moyennes
<b>Le niveau de confiance</b>	En général, 95 % – un niveau de confiance de 99 % demande un échantillon de 70 % plus grand, alors qu'un niveau de confiance de 90 % demande un échantillon 30 % plus petit			
<b>L'intervalle de confiance</b>	En général +/-5 % – utiliser un intervalle de +/- 3 % demande une taille d'échantillon 2,8 fois plus grande que pour un intervalle de +/-5%			
<b>La proportion de base</b>	Une proportion de 50 % demande un échantillon presque trois fois plus grand qu'une proportion de 10 % ou 90 %		Une proportion de base de 10 % peut nécessiter une taille d'échantillon au minimum 10 fois plus grande qu'une proportion de 50 %	
<b>La variabilité</b>		La taille de l'échantillon nécessaire est multipliée par quatre lorsque l'écart standard double		La taille de l'échantillon nécessaire est multipliée par quatre lorsque l'écart standard double
<b>La taille d'effet</b>			La taille de l'échantillon nécessaire est multipliée par quatre lorsque la taille d'effet diminue de moitié	
<b>La puissance</b>			Une puissance de 90 % demande un échantillon 33 % plus grand qu'une puissance de 80 % (à un niveau de confiance de 95 %)	

## 12.4 L'échantillon pour les méthodes combinées

Seules des méthodes quantitatives ou des recensements peuvent répondre de façon sûre à des questions portant sur le pourcentage d'une population ayant obtenu un bénéfice particulier ou connu un résultat particulier. Seules les méthodes qualitatives peuvent fournir de bonnes explications sur le pourquoi de ce qui s'est produit. On peut faire valoir que des méthodes quantitatives présentant une analyse des facteurs peuvent également fournir des explications, mais seulement pour les facteurs explicatifs qui sont explicitement inclus dans l'étude. La force des méthodes qualitatives est de permettre à l'évaluateur d'identifier les facteurs et les sujets qui apparaissent durant la recherche. C'est l'une des raisons pour lesquelles les méthodes qualitatives sont si utiles dans l'EAH où les contextes sont souvent chaotiques.

Les différentes forces des méthodes quantitatives et qualitatives pour signaler ce qui s'est produit et pourquoi expliquent pourquoi les meilleures EAH utilisent des méthodes combinées.

Les méthodes combinées requièrent à la fois des approches d'échantillonnage aléatoires et non aléatoires. Les évaluateurs humanitaires, qu'ils utilisent des méthodes quantitatives ou qualitatives, doivent expliquer leur stratégie d'échantillonnage et leurs choix. Ils doivent également expliquer tout biais potentiel ou limitations inhérentes à leurs choix.

# 13 / Les méthodes de terrain

Cette section se concentre sur les méthodes utilisées pour collecter des données autres en dehors des méthodes de revue documentaire. Ces dernières sont abordées dans la [Section 10 : Les méthodes de recherche documentaire](#).

Après une introduction générale sur les entretiens, différentes méthodes sont abordées. Pour chaque méthode, cette section examine :

1. Dans quelle mesure la méthode évoquée est utilisée dans l'EAH
2. En quoi elle consiste
3. Quelle sorte des données est collectée
4. Quel échantillonnage est généralement utilisé
5. Si elle est chère ou non
6. Combien peuvent être utilisées dans une petite EAH comprenant deux semaines de travail de terrain

Cette section propose également des astuces pour chaque méthode abordée.

L'analyse des données collectées est traitée dans la [Section 16 : L'analyse](#).

Le développement de la matrice d'évaluation est examiné dans la [Section 8 : La phase de démarrage](#).

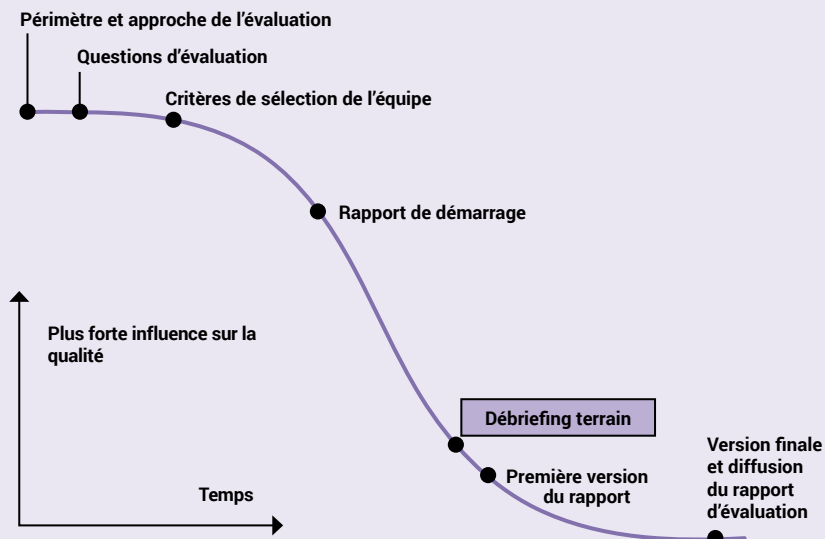


## A garder à l'esprit

Les questions d'évaluation déterminent les méthodes utilisées, mais des considérations pratiques de budget, de temps, d'éthique et de logistique peuvent empêcher que les méthodes les plus appropriées ne soient utilisées pour une question particulière.

Les évaluations peuvent utiliser diverses méthodes pour mesurer ce dont elles ont besoin. A titre d'exemple, l'évaluation IRC de la reconstruction du Comté de Lofa a utilisé des jeux comportementaux et des enquêtes pour tester si les communautés atteintes par le projet se comportaient différemment des communautés de contrôle : « Nous étions à la recherche de mesures directes d'effets comportementaux liés au programme CDR comme le révélaient les comportements observés dans le cadre des « jeux sur les biens publics » (Fearon *et al.*, 2008: 25).

## Point de contrôle de la qualité : Débriefing terrain



Le *débriefing* sur le terrain peut aider à identifier des lacunes importantes et influencer l'analyse des données collectées.

## 13.1 La technique de l'entretien

L'entretien est au cœur de nombreuses méthodes de recherche sociale et il peut être divisé en trois catégories :

1. **Les entretiens structurés.** Ils sont utilisés pour des enquêtes quantitatives et les intervieweurs respectent un protocole strict au niveau des questions posées.
2. **Les entretiens semi-structurés.** Très largement utilisés dans l'EAH. L'intervieweur interroge son interlocuteur pour extraire de ses réponses des données et dégager du sens.
3. **Les entretiens non structurés.** Dans ce cadre, l'entretien prend la forme d'une conversation ouverte bien qu'il puisse y avoir un thème central, comme le récit de sa vie. Cette technique est peu utilisée dans l'EAH.

Les entretiens sont des dialogues et pas seulement un transfert d'informations. Cela est aussi vrai de l'entretien formel d'enquête que de l'entretien qualitatif d'informateur clé (Suchman et Jordan, 1990). Les recherches réalisées sur les entretiens formels



montrent que les réponses varient non seulement en fonction de la manière dont la question est posée mais aussi du sexe de l'intervieweur, de son appartenance ethnique, de son âge, de sa classe sociale « apparente » et de sa taille corporelle (Dohrenwend *et al.*, 1968 ; Eisinga *et al.*, 2011 ; Flores-Macias et Lawson, 2008 ; Kane et Macaulay, 1993 ; Riessman, 1977 ; Suchman et Jordan, 1990 ; Webster, 1996).



**A garder à l'esprit**

Dans certains contextes, le fait d'interroger des personnes peut les mettre en danger. Ne réalisez pas d'entretiens dans de telles circonstances à moins que les personnes interrogées ne soient pleinement conscientes des risques et les acceptent. Même si elles peuvent encourir un risque à cause de l'entretien, elles ont aussi le droit de faire entendre leur voix (voir [Section 14 : Impliquer la population affectée dans votre évaluation](#)).

**Les entretiens structurés**

Les entretiens structures sont généralement conduits dans le cadre d'enquêtes quantitatives. Afin de minimiser les biais et les erreurs qui en résultent dans les calculs, les entretiens sont réalisés de façon standardisée. Les intervieweurs ne sortent pas du cadre des questions définies.

Les intervieweurs portent généralement le nom d'enquêteurs, et ils peuvent être fournis par une société d'études de marché ou être des étudiants issus de l'université. Les enquêteurs n'ont pas besoin des mêmes niveaux de compétences et de compréhension du contexte que les intervieweurs réalisant des entretiens qualitatifs ; de ce fait, ils ont tendance à coûter moins chers.

On demande aux enquêteurs de poser les questions dans un ordre défini et de ne pas expliquer les questions de façon à éviter de biaiser les réponses puisque les autres personnes interrogées auraient pu répondre différemment si elles avaient reçu la même explication. Pour souligner son caractère formel, le questionnaire utilisé est généralement désigné comme l'outil d'enquête.



**Définition : L'outil d'enquête**

L'outil d'enquête est le questionnaire utilisé par l'intervieweur durant les entretiens d'enquêtes formelles.

Un outil d'enquête peut présenter un grand nombre de questions et être organisé en sections de manière à examiner différents concepts comme le *leadership* ou les compétences. L'outil d'enquête destiné aux enseignants dans le cadre de l'étude NRC sur l'éducation en Palestine (Shah, 2014) comprenait 91 questions qui mesureraient dix concepts sous-jacents. Une partie standard de l'outil d'enquête teste dans quelle mesure chaque question est liée aux autres pour ce concept (la fiabilité interne).

Les questions sont extraites de l'outil d'enquête. Il est essentiel de tester l'outil d'enquête à l'avance de manière à identifier et reformuler les questions qui amènent des interrogations ou des réponses incertaines de la part des personnes interrogées (voir [l'exemple de bonnes pratiques p. 251](#) du WVI portant sur la traduction et la traduction en sens inverse).

Même si les entretiens structurés peuvent comprendre certaines questions ouvertes, celles-ci posent deux problèmes :

- La réponse doit être codée d'une manière ou d'une autre de façon à pouvoir être analysée, ce qui peut être onéreux et chronophage.
- La réponse à la question ouverte peut soulever d'autres questions que l'enquêteur n'a pas le droit de poser ou qu'il n'a pas les moyens de suivre.

### Les entretiens semi-structurés

Les entretiens semi-structurés sont à la base de diverses méthodes largement utilisées dans l'EAH. Dans les entretiens semi-structurés, l'intervieweur utilise un guide d'entretien ou une liste de grands thèmes à couvrir et pose des questions à l'intérieur de ce cadre. Dans cette approche, l'intervieweur revient sur des réponses antérieures et pose des questions d'approfondissement pour découvrir des données et dégager du sens.

Les questions d'approfondissement<sup>4</sup> sont des questions qui cherchent à :

- **Clarifier les réponses** – lorsqu'un intervieweur n'est pas sûr de ce que la personne interrogée a voulu dire
- **Obtenir plus de détails** – l'utilisation la plus courante des questions d'approfondissement dans le cadre d'entretiens d'EAH
- **Connaître l'analyse de la personne interrogée** – peut être très utile pour comprendre pourquoi les individus ont agi d'une façon particulière
- **En savoir plus sur les variations et les cas contrefactuels** – utiles pour faire ressortir les arguments qui se trouvent derrière les décisions prises.

Les questions d'approfondissement utilisées dépendront des questions d'évaluation et du périmètre de l'évaluation.

#### Clarifier les questions d'approfondissement

Cela peut comprendre :

- Est-ce que je vous ai bien compris quand vous dites que...? C'est souvent la question d'approfondissement qui clarifie le mieux car elle ne sous-entend pas que la personne interrogée n'a pas été claire dans sa réponse précédente.
- Quand vous dites... que voulez-vous dire exactement ?
- Vous avez utilisé le terme ... Ai-je raison de penser que vous renvoyez à ...?

### Les questions d'approfondissement pour préciser des points de détail

Cela peut comprendre :

- Pouvez-vous me donner un exemple ?
- Comment gérez-vous cela ?
- Quelle réponse avez-vous reçue par rapport à cela ?
- Pouvez-vous m'en dire plus à propos de votre propre situation ?

### Les questions d'approfondissement pour préciser des points d'analyse

Cela peut comprendre :

- Comment qualifieriez-vous ce qui s'est passé ?
- Qu'est-ce que vous retenez particulièrement de cela ?
- Qu'est-ce qui était important par rapport à cela ?
- Quels étaient les résultats auxquels vous vous attendiez ?
- Comment vous êtes-vous sentis par rapport à cela ?

### Les questions d'approfondissement pour préciser des variations et des points contradictoires

Cela peut comprendre :

- Votre approche a-t-elle changé au fil du temps ? En quoi a-t-elle changé ? Qu'est-ce qui a motivé ce changement ?
- Procéderiez-vous de la même manière la prochaine fois ?
- Je note que vous faites ... mais l'organisation Y fait ... sur la même thématique. Quels sont les avantages de votre approche par rapports aux leurs, et inversement ?
- Certaines des personnes à qui j'ai parlé ont dit ... Quel est votre avis sur ce point ?

## Les entretiens non structurés

Ces entretiens peuvent aller des entretiens totalement non structurés comme ceux utilisés dans le cadre de recherches ethnographiques jusqu'aux entretiens thématiques où l'intervieweur introduit un thème, comme le récit de la vie de la personne interrogée, et la laisse poursuivre.

Les entretiens non structurés n'ont pas été beaucoup utilisés dans l'EAH, mais certaines méthodes d'entretien, comme les discussions en *focus group*, comprennent un élément propre à l'entretien non structuré. Un exemple d'utilisation de l'entretien thématique non structuré était l'évaluation Disaster Assistance Committee (DAC) des inondations qui ont eu lieu en 2000 au Mozambique. Dans ce cadre, un évaluateur a conduit de longs entretiens avec des membres de la population affectée pour se faire une idée de leur vécu par rapport à la réponse humanitaire (Cosgrave *et al.*, 2001).

## Les biais propres aux entretiens

Il est évident que tous les intervieweurs et les personnes interrogées ont des intérêts. Certains intervieweurs travaillant pour des organisations peuvent vouloir que leur travail soit largement connu, d'autres peuvent vouloir masquer des faiblesses. La population affectée peut souhaiter que l'assistance se prolonge, ou préférer qu'elle s'arrête. L'évaluateur doit rester conscient du potentiel de tels biais, et du fait que les entretiens sont des processus construits à partir des valeurs et des intérêts divergents de l'intervieweur et de la personne interrogée. Le chapitre consacré à l'analyse s'intéresse à la manière de traiter les biais des éléments de preuve tirés d'entretiens.

Comme avec toutes les méthodes qualitatives ou combinées, la triangulation est la technique la plus efficace pour maîtriser les biais. La triangulation n'est jamais aussi puissante que lorsque la sélection raisonnée des personnes interrogées est aussi inclusive que possible. La triangulation – qui renvoie aux biais des entretiens et aux conflits de données – est abordée plus loin dans la [Section 13 : Les méthodes de terrain](#).



### Astuces :

Lorsque l'on réalise des entretiens avec la population affectée, le timing des entretiens a une grande influence sur le profil des personnes interrogées qui seront disponibles. Les évaluateurs doivent être conscients du type de personnes qu'ils excluent des entretiens en les organisant à certains moments précis de la journée.

Voir [Section 11 : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation sur les biais en général](#).

## 13.2 Les interprètes

Idéalement, les évaluateurs devraient être capables de parler la langue des personnes qu'elles interrogent. Si ce n'est pas le cas, ils doivent travailler avec des interprètes pour certains types d'entretiens. La présence d'un interprète fait toujours ressortir des problèmes au sein de l'évaluation. Ce n'est pas quelque chose qui va sans heurts et Temple *et al.* (2006) affirment que les entretiens qui ont recours à un interprète sont plus semblables à de l'analyse de données secondaires qu'à de la recherche primaire.

Même si des interprètes peuvent être utilisés pour des entretiens semi-structurés et des entretiens de groupe, il est presque impossible d'utiliser des interprètes pour conduire des entretiens structurés efficaces ou des discussions en *focus group*. Les entretiens d'enquêtes prennent déjà un temps considérable et l'ajout d'un interprète les rend encore plus longs. Dans le cadre de *focus groups*, les mécaniques d'interprétariat interfèrent avec l'idée d'échanger des propos librement.

Pour les *focus groups*, Steward *et al.* (2007: 112) note que :  
 Les modérateurs doivent avoir pour langue maternelle la langue dans laquelle le groupe s'entretient. Même la personne la plus à l'aise dans ce qui serait sa seconde langue aura des difficultés à formuler certaines questions et suivre les réponses. Les interprètes créent des problèmes encore plus importants et doivent être évités lorsque cela est possible. Il arrive souvent que les perceptions et les opinions de l'interprète colorent la traduction.

Devoir passer par un interprète augmente le temps requis pour passer en revue une série de questions. Cela peut parfois être utile lors d'entretiens qualitatifs car cela donne plus de temps à l'évaluateur pour réfléchir aux réponses associées aux questions précédentes tout en formulant des questions d'approfondissement. Les évaluateurs doivent faire tout leur possible pour engager des interprètes qualifiés. Les interprètes compétents peuvent également apporter une plus-value en faisant ressortir les points contextuels des discussions une fois l'entretien terminé, nuances que l'évaluateur aurait pu négliger sans cela.



**Astuce :**

Engagez les meilleurs interprètes possibles par rapport à votre budget. Les interprètes professionnels justifient leur salaire parce qu'ils sont moins susceptibles de déformer les questions et les réponses. Par conséquent, accordez-vous cette sécurité au niveau du budget.

Rappelez-vous qu'il faut du temps pour recruter des interprètes. Voyage  
 Les dispositions relatives au voyage et au logement doivent permettre la présence des interprètes.

Lors d'une évaluation en Afghanistan, on a demandé aux personnes âgées d'une communauté quel était l'impact des Talibans sur leur accès aux services gouvernementaux. Le membre de l'organisation jouant le rôle d'interprète a dit à l'équipe que les personnes âgées se plaignaient que les actions des Talibans avaient réduit leur accès à ces services. Par la suite, l'assistant de l'équipe d'évaluation leur a rapporté que les anciens avaient dit en réalité qu'ils *étaient* les Talibans et que le gouvernement refusait de leur fournir les services à moins qu'ils ne rendent leurs armes.



**Astuce :**

Évitez si possible d'utiliser du personnel de l'organisation comme interprète.

Même si c'est loin d'être l'idéal, il n'est pas toujours possible d'éviter d'utiliser le personnel de l'organisation comme interprète parce que les interprètes professionnels ne sont pas disponibles pour les langues pertinentes ou encore pour des raisons de sécurité ou de restrictions d'accès.



Lors de l'évaluation de la réponse CARE à la crise humanitaire de 2011-2012 au Sahel (Gubbels et Bousquet, 2013: 18), l'équipe a remarqué que, tout en conduisant des discussions de *focus groups*, elle était : obligée de prendre un des animateurs de CARE au Tchad comme interprète. C'était une source potentielle de biais. Toutefois, en utilisant des questions d'approfondissement lorsque c'était nécessaire, et en évaluant le schéma général des réponses au sein de 6 villages, l'équipe d'évaluation pense avoir évité le problème du biais positif.

**Astuce :**

Recoupez les interprétations des personnes qui ont un intérêt manifeste. Si vous n'avez aucune connaissance de la langue et que vous devez vous fier à un membre de l'organisation pour tenir le rôle d'interprète, il est bon d'avoir un assistant qui comprenne la langue et puisse avertir l'équipe en cas d'inexactitudes graves. Vous pouvez également enregistrer de façon visible les échanges pour les analyser plus tard et ainsi encourager une interprétation précise.

## 13.3 Les méthodes d'entretien qualitatif

### Les entretiens avec des informateurs clés

Les entretiens avec des informateurs clés constituent la colonne vertébrale de l'EAH.

**Éléments clés :**

- Ils utilisent la technique semi-structurée et un guide d'entretien.
- Durent généralement 45 à 60 minutes.
- Fournissent des données qualitatives, conviennent mieux à ce qui relève de l'analyse et de la compréhension.
- Approprié pour répondre à des questions portant sur le comment et le pourquoi des choses qui se sont produites.
- Inadapté à la collecte de données quantitatives.
- Demandent des intervieweurs expérimentés qui peuvent poser des questions d'approfondissement pour dégager du sens.
- Généralement consignés en prenant des notes, mais peuvent être enregistrés avec un appareil numérique.
- L'échantillonnage est en général raisonné. Les personnes interrogées sont sélectionnées parce que l'on s'attend à ce qu'elles soient les mieux à même de répondre à des questions d'évaluation.
- Méthode rapide et relativement peu onéreuse.
- Deux semaines de travail de terrain peuvent permettre de réaliser 25 à 50 entretiens avec des informateurs clés.

- En général, pas plus de quatre ou cinq par jour.
- La saturation théorique (un plus grand nombre d'entretiens n'apportent pas de nouvelles informations) propre à un type particulier de personnes interrogées se produit entre 6 et 12 entretiens d'un type particulier (Guest *et al.*, 2011) mais il peut être nécessaire d'en faire plus pour des questions de crédibilité.



### **A garder à l'esprit**

De nombreuses questions éthiques sont posées par le fait que les entretiens ne doivent pas mettre en danger les personnes interrogées et que la confidentialité doit être respectée.



### **Astuces :**

Vous pouvez légèrement accélérer le déroulement des entretiens en transmettant à l'avance aux personnes interrogées des détails sur l'objectif et les règles de base.

Les entretiens avec des informateurs clés sont généralement réalisés dans un espace privé de manière à ce que d'autres personnes n'entendent pas par hasard ou ne se joignent pas à la conversation. Vous pouvez également utiliser des espaces publics respectant des normes en matière de vie privée, comme un café.

L'évaluateur est souvent perçu comme un représentant de la communauté de l'aide. Dans de nombreuses cultures, il est inconvenant de critiquer quelque chose auprès d'une personne censée représenter l'organisation qui l'a fournie. Dans de telles cultures, il peut être plus utile de demander aux personnes comment leurs voisins jugent l'assistance fournie.



## Pour approfondir : Le guide d'entretien

Le guide d'entretien expose la tendance générale des questions, débutant avec un rappel signifiant à la personne interrogée l'objectif de l'évaluation et les règles de base de la conversation (en général, les commentaires émis durant l'entretien ne sont pas attribuables, que ce soit de manière directe ou indirecte).

Il est bon de commencer en demandant à la personne interrogée de décrire son rôle – presque tout le monde est ravi de parler de soi, même si l'on est timide. Toutes les questions du guide d'entretien ne seront pas pertinentes pour toutes les personnes interrogées et l'intervieweur ne doit pas poser de questions hors de propos.

Les questions doivent être organisées de façon à établir une relation – en reportant les questions plus litigieuses vers la fin de l'entretien.

Le guide d'entretien doit comprendre au plus 20 à 25 questions. S'il en contient plus, il sera nécessaire de les prioriser ou de préparer un guide séparé pour les différents types de personnes interrogées.

Les questions de clôture doivent comprendre une question sur ce que la personne interrogée a appris ou ce qui a été renforcé par la réponse humanitaire. C'est souvent une bonne question si les questions précédentes ne l'ont pas fait ressortir. Demander à la personne interrogée si elle a été surprise par l'omission de certains sujets peut contribuer à mieux comprendre le but de l'évaluation et fournit parfois de nouvelles pistes.

Les questions de clôture doivent également demander aux personnes interrogées de suggérer d'autres personnes et groupes de personnes pour des entretiens. Si nécessaire, demandez à la personne interrogée si elle serait disponible pour des questions complémentaires un peu plus tard dans le processus d'évaluation.



### Astuces :

Évitez les questions qui permettent d'obtenir la « bonne » réponse. Par exemple, si vous voulez savoir si l'équipe de mise en œuvre a été sensible aux questions de genre durant l'intervention, ne demandez pas : « Avez-vous fait attention aux questions de genre ? » car la réponse sera toujours « oui ». Demandez plutôt : « Qui - des femmes ou des hommes – a le plus bénéficié de l'intervention et pourquoi ? », « Quelles différences notez-vous entre les effets produits sur les hommes et sur les femmes ? »



## Les entretiens avec des ménages

Ils sont couramment utilisés dans le cadre d'enquêtes quantitatives, mais ils peuvent également être utilisés pour collecter des données qualitatives.

### Éléments clés :

- Ils peuvent utiliser un outil d'enquête formel ou un guide d'entretien ou encore une combinaison des deux.
- Peuvent durer de 20 à 60 minutes selon le niveau de complexité.
- Peuvent recueillir des données qualitatives et quantitatives, selon le type d'entretien.
- Approprié pour répondre à des questions sur l'impact de l'assistance au niveau des foyers.
- Nécessitent des intervieweurs qui comprennent bien le contexte et peuvent relever des indices à partir de l'observation de manière à trianguler ce qu'on leur a dit.
- Généralement consignés dans un formulaire d'enquête, mais peuvent être enregistrés avec un appareil numérique.
- Le lieu de l'entretien est choisi de façon raisonnée, mais les ménages peuvent être sélectionnés de façon aléatoire (pour les études à grand nombre de données) ou raisonnée (pour les études à petit nombre de données).
- Cette méthode est modérément coûteuse, à moins qu'elle ne fasse partie d'un système de suivi régulier.
- Généralement réalisés comme exercice distinct si l'on veut qu'ils soient représentatifs.
- Le nombre d'entretiens par jour dépend du niveau de complexité et du temps de transport.

Lorsque les entretiens avec des ménages servent à recueillir des données qualitatives, ils couvrent en général seulement quelques sujets et sont bien plus courts qu'un entretien complet avec un informateur clé. Leur principal avantage sur les entretiens avec des groupes est que le ménage est un espace privé et qu'il est de ce fait possible d'aborder des questions plus sensibles.



#### Astuce :

Les entretiens avec des ménages peuvent fournir des données intéressantes sur des cas révélateurs qui font ressortir des points particuliers. De tels cas exemplaires peuvent apporter une dimension humaine aux lecteurs et améliorer le rapport d'évaluation.

## Les entretiens avec des groupes

### Éléments clés :

- Ils utilisent la technique semi-structurée et un guide d'entretien tronqué car cinq à dix questions tout au plus peuvent être posées.
- Durent généralement 30 à 45 minutes, mais ils peuvent être plus longs si les personnes interrogées sont intéressées par les sujets traités.
- Recueillent des données qualitatives, conviennent mieux à ce qui relève de l'analyse et de la compréhension.
- Approprié pour répondre à des questions portant sur les perceptions communautaires de l'aide.
- Inadapté à la collecte de données quantitatives.
- Nécessitent des intervieweurs expérimentés qui peuvent connaître bien la communauté et poser des questions d'approfondissement pour dégager du sens.
- Généralement consignés en prenant des notes, mais peuvent être enregistrés avec un appareil numérique.
- Le lieu de l'entretien est choisi de façon raisonnée, et les premières personnes interrogées peuvent également être sélectionnées de façon raisonnée. Toutefois, comme ces entretiens sont généralement menés dans des espaces publics, d'autres personnes peuvent se joindre à la conversation.
- Méthode rapide et relativement peu onéreuse.
- Deux semaines de travail de terrain peuvent permettre de réaliser 5 à 20 entretiens avec des groupes.
- En général, pas plus de deux à trois par jour.
- Pour ce type d'entretiens, il n'existe pas d'estimations standards relatives à la saturation théorique mais on peut penser qu'il s'agit de deux ou trois pour les *focus groups* et de six à douze pour les entretiens avec des informateurs clés.
- Ont tendance à être dominés par certains individus, en particulier dans les sociétés hiérarchisées.



#### **A garder à l'esprit**

Peuvent soulever de nombreuses questions éthiques du fait d'être conduits en public. En outre, le fait de parler en public peut mettre en danger les personnes interrogées.



#### Astuce :

- Gérez le groupe. Si la réunion se tient à l'extérieur, ne restez pas debout mais essayez par terre si possible. Cela obligera le groupe à s'asseoir et vous éviterez ainsi que les gens ne se poussent ou ne se bousculent.
- Renseignez-vous sur les normes sociales locales et respectez-les autant que possible.
- Dirigez vos questions vers les participants qui ne sont pas les plus bavards.  
Demandez combien de personnes sont d'accord avec un point de vue particulier.

## Les réunions formelles avec les structures communautaires

Une variante de la réunion de groupe couramment pratiquée dans certaines régions d'Asie est la réunion organisée par des structures communautaires. Elles peuvent être assez formelles avec une table et une personne désignée pour conduire la réunion. De telles réunions tendent à être dominées par les chefs communautaires, ce qui peut compliquer la collecte d'informations provenant de l'ensemble de la communauté.

La meilleure approche consiste à utiliser des questions dont vous espérez qu'elles provoqueront des réponses différentes au sein du groupe. Les questions peuvent par exemple porter sur : qui a été le plus affecté, qui a le plus bénéficié de l'intervention entre les hommes et les femmes, quel type d'assistance a été le plus utile la première semaine. Les premières réponses spontanées seront peut-être difficiles à obtenir mais de telles réunions peuvent fournir des données utiles.



### Exemple de bonnes pratiques : Développer une politique de résolution des conflits

Dans le cadre d'une évaluation UNHCR portant sur la transversalisation des approches âge, genre et diversité (Thomas et Beck, 2010), l'étude Colombie a utilisé une approche extrêmement participative. Quatre communautés de différentes parties du pays ont été sélectionnées pour participer à l'évaluation tout au long des étapes suivantes (Mendoza et Thomas, 2009) :

- Un atelier avec des membres de la communauté dans le but d'élaborer une chronologie des événements ayant mené au diagnostic participatif de l'UNHCR (un outil clé dans le processus de transversalisation) et le plan d'action qui en a résulté, pour mesurer la sensibilisation au plan d'action et juger si le diagnostic participatif avait généré des changements dans la communauté.
- Des réunions avec des sous-groupes de femmes, d'hommes, d'adolescent(e)s, d'enfants, de personnes âgées et de personnes handicapées pour vérifier si ces sous-groupes avaient des perceptions différentes des résultats du diagnostic participatif et du plan d'action.

### Les discussions en *focus group*

Même si le terme *focus group* s'applique très largement aux entretiens de groupe, il est plutôt réservé aux entretiens structurés.

#### Éléments clés :

- Durent en général deux heures.
- Six à huit participants en général.
- S'appuient sur un guide des sujets comprenant trois à cinq sujets de discussion.
- Très efficaces avec le personnel des organisations partenaires, mais peuvent également être utilisées avec la population affectée.
- Génèrent des données qualitatives, intéressant pour explorer les points de vue de groupes spécifiques (parents célibataires, jeunes, etc.)
- Nécessitent un espace confortable et contrôlé.
- Nécessitent un facilitateur et un preneur de notes.
- Généralement enregistrés en audio ou vidéo (voir [Section 14 : Impliquer la population affectée dans votre évaluation](#)).
- Des échantillons peuvent être établis de façon aléatoire au sein des groupes dans le but de former le *focus group*. Il ne doit pas y avoir de différences de statut ou d'autre nature à l'intérieur du groupe.
- Morgan (1997) note que peu de thèmes nouveaux émergent après le troisième *focus group*.
- Coût moyen.
- Une évaluation d'une durée de deux semaines comprend généralement moins de dix discussions en *focus group*.

- Dans certains pays, les participants aux *focus groups* peuvent être rémunérés. Ce n'est pas la pratique dans l'EAH, mais il faut au minimum offrir des boissons dans le cadre d'un engagement relativement long.

Les ouvrages de référence en matière de focus group sont ceux de Krueger et Casey (2009) et de Steward *et al.* (2007).



**Astuce :**

Interagissez avec le groupe avant la discussion de manière à repérer les plus bavards. Prenez une collation ensemble avant le début de la discussion en focus group de façon à identifier les personnes qui dominent les autres. Placez-les à côté du facilitateur dans le but de réduire les contacts oculaires. Placez les plus calmes en face du facilitateur pour qu'il les regarde plus souvent.

En analysant l'enregistrement du *focus group*, vous vous intéresserez aux mots, au contexte, aux changements de points de de vue, à la fréquence des commentaires particuliers, au nombre de discussions, à leur intensité et leur spécificité, ainsi qu'au contenu.



**Exemple de bonnes pratiques : L'utilisation que World Vision fait des discussions en focus group (DFG)**

« Pour les DFG, chaque équipe a en général un chef d'équipe, même si parfois nous ne sommes pas en mesure de procéder ainsi. Dans ce cas-là, nous utilisons la structure régionale (par exemple : deux ou trois équipes DFG dans une région, avec un chef d'équipe DFG qui les encadre toutes). Nos équipes DFG comprennent entre quatre et six personnes : le chef d'équipe, un traducteur, deux preneurs de notes, un observateur et un animateur. L'observateur et l'animateur ne sont pas du même sexe ; ils changent de rôle pour que l'animateur soit toujours du même sexe que le DFG ; l'observateur, le chef d'équipe et le traducteur peuvent également prendre des notes si la discussion est très animée. »

Source : Kathy Duryee, WVI, communication personnelle, 2014

## 13.4 Les méthodes de sondage

### L'utilisation de sondages

Les sondages, en particulier les sondages formels en face-à-face, sont considérés comme très officiels. Cela s'explique en partie parce que de nombreux lecteurs n'ont pas une bonne compréhension des statistiques. Prenons l'exemple suivant :



#### **Un sondage politique indique une baisse de 2 % du soutien à un parti... Qu'est-ce que cela signifie ?**

De nombreux sondages politiques utilisent un niveau de confiance de 95 % et une marge d'erreur de +/- 3%. Le sondage signale une baisse de 35 % à 33 % du soutien au parti au pouvoir depuis le dernier sondage.

#### **Hypothèse courante :**

33 % de la population soutient désormais le parti au pouvoir.

- **Situation réelle :**

33 % des personnes échantillonnées ont soutenu le parti au pouvoir.

#### **Hypothèse courante :**

Il y a 95 % de chance que 33 % de la population soutiennent le parti au pouvoir.

- **Situation réelle :**

Il y a 95 % de chance qu'entre 30 % et 36 % (33 % +/- 3%) de la population soutienne le parti au pouvoir. Il y a également 5 % de chance que ce soutien soit inférieur à 30 % ou supérieur à 36 %.

#### **Hypothèse courante :**

Le soutien au parti au pouvoir a chuté.

- **Situation réelle :**

Auparavant, le niveau de soutien de la population au parti au pouvoir se situait avec un niveau de probabilité de 95 % entre 32 % et 38 % (35 % +/- 3 %) et il se situe à présent entre 30 % et 36 %. Il est même concevable que le soutien populaire ait augmenté de 32 % à 36 %. Toutefois, tous les résultats ne sont pas aussi probables les uns que les autres car les mesures de l'échantillon se trouvent en général autour de la moyenne de la population. *In fine*, il est donc plus probable que la moyenne de la population soit plus proche de la moyenne de l'échantillon que de ses valeurs inférieures et supérieures. C'est pourquoi les sondages indiquent rarement de grandes variations.

Il est très important d'avoir des connaissances de base en matière de statistiques si l'on veut comprendre les rapports de sondage. Alors que les équipes peuvent rendre compte du chevauchement de marges d'erreur entre des sondages avant/après, elles se concentrent sur les changements de la valeur moyenne de la variable d'intérêt plutôt que sur la possible conclusion qu'il n'y ait eu aucun changement.

## Les enquêtes nutritionnelles

Les enquêtes nutritionnelles sont largement utilisées pour suivre les programmes nutritionnels. Les évaluations tendent à utiliser des données d'enquêtes plutôt qu'à mener leur propre enquête, comme dans le cas de l'évaluation de l'aide alimentaire au Rwanda pour les réfugiés de longue durée (Sutter *et al.*, 2012). De telles enquêtes sont assez techniques et ne sont pas analysées ici.

## Les enquêtes formelles en face-à-face

Ces enquêtes peuvent utiliser les individus comme point d'échantillonnage mais dans l'EAH, il est plus fréquent de prendre les ménages comme point d'échantillonnage. L'unité d'analyse peut être le ménage ou un membre du ménage, sélectionné à partir d'un formulaire pseudo-aléatoire comme la date d'anniversaire la plus récente, la grille de Kish<sup>5</sup> ou un échantillonnage par quotas (hommes et femmes tour à tour de manière à assurer un équilibre entre les sexes). Malgré cela, parler avec quelqu'un d'autre que l'informateur initial peut parfois être difficile pour des raisons culturelles (McBurney, 1988).

### Éléments clés :

- Utilisent un outil d'enquête que les intervieweurs suivent à la lettre.
- La durée de chaque entretien dépend de la longueur et de la complexité de l'outil d'enquête.
- Recueillent des informations quantitatives – intéressants pour répondre à des questions de type « Combien de... ». Peuvent également poser des questions qualitatives fermées car les questions ouvertes sont très exigeantes en termes de codage et d'analyse.
- Les enquêteurs n'ont généralement qu'une formation de quelques jours.
- Les données peuvent être enregistrées dans un formulaire d'enquête ou à l'aide d'un appareil numérique. L'enregistrement électronique est de loin préférable.
- Les personnes interrogées sont sélectionnées de façon aléatoire de manière à ce que les résultats de l'enquête puissent être généralisés à l'ensemble de la population. Chaque écart commis par rapport à un véritable échantillonnage aléatoire (par exemple l'échantillonnage pseudo-aléatoire ou par grappes) augmente la taille de l'échantillon nécessaire pour aboutir à des résultats valides sur le plan statistique.
- En général, plusieurs centaines d'entretiens sont nécessaires, mais cela peut aller jusqu'à plusieurs milliers.
- Onéreux car elles demandent de tester et de valider l'enquête et qu'un grand nombre d'entretiens est nécessaire. Son utilisation dans l'EAH tend à être limitée aux évaluations de grande échelle disposant d'un important budget en conséquence.

Les enquêtes en face-à-face peuvent s'avérer un moyen très efficace pour définir l'importance de différents sujets, notamment l'ampleur de l'impact d'une intervention.

Les enquêtes portant sur la population affectée sont très efficaces pour révéler des perceptions de l'aide, des limites de couverture, de respect des délais, etc. Si vous pouvez vous le permettre et que vous n'avez pas déjà engagé la population dans d'autres méthodes, il est pertinent de réaliser une enquête auprès de la population affectée.

La taille de l'échantillon nécessaire varie en fonction de divers facteurs, notamment de si vous voulez parler de prévalence seulement dans un groupe ou si vous voulez comparer différents groupes ou encore comparer le même groupe dans la durée. Voir Section 11 : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation pour des conseils sur l'échantillonnage.

L'ouvrage de Fowler (2009) constitue une bonne introduction succincte aux enquêtes.

Les enquêtes formelles deviennent plus courantes dans l'EAH, en particulier dans le cadre de grandes évaluations dotées d'un budget important.

## Exemples d'enquêtes dans l'EAH

Voici quelques exemples d'utilisation d'échantillons formels dans l'EAH.

- Les enquêtes réalisées auprès de la population affectée après le Tsunami de 2004 dans l'océan Indien, le tremblement de terre au Pakistan en 2006 et le séisme de Yogyakarta en 2007 (Bliss et Campbell, 2007a, 2007b ; Bliss *et al.*, 2006 ; Fritz Institute, 2005a, 2005b, 2005c).
- L'étude LRRD de la Tsunami Evaluation Coalition (Christoplos, 2006) qui a interrogé 1.227 personnes en Indonésie et 915 au Sri Lanka.
- Le suivi de cette étude LRRD (Brusset *et al.*, 2009) auprès de 1.178 personnes en Indonésie et de 965 au Sri Lanka.
- L'évaluation inter-organisations de l'appui aux réfugiés dans le cadre de déplacements prolongés comportait des enquêtes dans chacun des pays retenus comme études de cas, notamment l'Ethiopie avec une taille d'échantillon de 1.081 (Sutter *et al.*, 2011), le Rwanda avec une taille d'échantillon de 1.200 (Sutter *et al.*, 2012) et le Bangladesh avec une taille d'échantillon de 1.069 (Nielsen *et al.*, 2012).



Une fois que les données de l'enquête ont été collectées, elles doivent être saisies et nettoyées. Cela peut être onéreux et chronophage, en particulier si l'enquête contient des questions ouvertes.



**Astuce :**

Utilisez du matériel électronique pour collecter les réponses des enquêtes. Cela éliminera les coûts de saisie de données et une grande partie du coût de nettoyage, car le formulaire d'enquête peut être développé de manière à rejeter les combinaisons de réponses non valables (comme un informateur âgé de seize ans qui aurait douze enfants). De plus, les données collectées peuvent être envoyées sur la base de données principale sans nécessiter d'autres saisies. Cela peut également accélérer le processus de l'enquête.

Des outils numériques peuvent être utilisés dans le cadre d'enquêtes de diagnostic et d'enquêtes d'évaluation. Le mécanisme de diagnostic rapide multi-Clusters au Pakistan a utilisé des assistants numériques personnels (PDA) pour recueillir des données de diagnostic (McRAM Team, 2009). Shirima *et al.* donnent un exemple de l'utilisation de PDA dans une enquête auprès de 21.000 ménages en Tanzanie et fournit quelques éclaircissements sur les questions pratiques que cela a impliqué (Shirima *et al.*, 2007). L'état actuel de la recherche laisse entendre que l'utilisation de PDA pourrait réduire le poids de la logistique, le coût et le taux d'erreurs de la collecte des données (Seebregts *et al.*, 2009).

Les téléphones portables sont de plus en plus utilisés par les enquêteurs afin de saisir directement les données. Ils présentent l'avantage de permettre le téléchargement immédiat des données, et ainsi de prévenir le risque de perte des données si le téléphone est endommagé ou perdu. Leur utilisation permet également un contrôle qualité en temps réel et une supervision de l'enquêteur de manière à réduire le risque de fabrication de fausses données (Tomlinson *et al.*, 2009). Les téléphones portables offrent des avantages similaires aux PDA pour ce qui est de réduire les erreurs d'enregistrement et de saisie (Zhang *et al.*, 2012). Plusieurs suites de logiciels comme SurveyToGo ou Snap Surveys peuvent être utilisées pour mettre en place des enquêtes s'appuyant sur les téléphones portables. Ces différents produits ont des modèles tarifaires distincts, et varient en fonction des appareils pris en charge. Ce secteur se développant très rapidement, il est important de vérifier les produits actuellement disponibles et les plus adaptés.

Les enquêtes posent un certain nombre de problèmes logistiques dans le cadre de l'EAH :

- **Absence d'un cadre d'échantillonnage pratique.** Un cadre d'échantillonnage est une liste plus ou moins précise de la population à partir de laquelle l'échantillon doit être établi. Quand il n'y a pas d'autre alternative, il convient d'utiliser les échantillons par grappes en sachant que cela augmente le nombre de personnes interrogées nécessaires par un facteur de deux (nutrition) ou de quatre (mortalité). Voir la [Section 12 : L'échantillonnage](#) pour une analyse de l'effet de conception.
- **Temps.** Développer et tester l'outil d'enquête peut prendre du temps. Faire les traductions, former les enquêteurs et réaliser l'enquête sont chronophages, et les enquêtes prennent souvent plus de temps que d'autres méthodes d'EAH.
- **Absence d'enquêtes sur les seuils de référence.** Les enquêtes sont plus utiles lorsqu'il existe un seuil de référence à partir duquel comparer les résultats. C'est rarement le cas dans les EAH mais la situation est en cours d'amélioration parallèlement à l'amélioration des statistiques nationales dans de nombreux pays en voie de développement.
- **Absence d'évaluateurs présentant les qualifications requises.** La plupart des évaluateurs humanitaires viennent d'une formation qualitative plutôt que quantitative, et n'ont pas beaucoup d'expérience dans les enquêtes quantitatives.

Concevoir un outil d'enquête est une tâche complexe. Les questions de l'enquête, bien que plus simples par certains aspects que les questions d'entretiens semi-structurés, sont plus complexes à d'autres égards, notamment parce que l'intervieweur ne peut pas ajuster la question et poser des questions d'approfondissement. Les enquêtes structurées en face-à-face partent du principe que les questions seront posées de la même manière à chaque personne interrogée. Les questions d'une enquête doivent être conçues avec précaution puis testées avant d'être utilisées car elles ne peuvent pas être ajustées une fois que l'enquête est en cours.

On teste habituellement les questions d'une enquête en réalisant une enquête pilote. La manière dont les questions sont posées peut avoir un impact important sur les réponses et une enquête pilote peut mettre en évidence certains de ces points. La monographie de Fowler (1995) sur les questions d'enquête est une ressource utile.



**Astuce :**

Si l'enquête doit être traduite, faites retraduire la traduction dans la langue originale par un autre traducteur.



### Exemple de bonnes pratiques : L'approche de World Vision International en matière de collecte de données

Une fois que la version anglaise des outils de collecte de données est prête, le processus suivant doit être respecté.

1. Un traducteur originaire de la région où l'enquête sera organisée (souvent du personnel local de WVI, ou parfois un consultant) est engagé pour les traduire dans la langue ou le dialecte local.
2. Les outils sont retraduits en sens inverse (en anglais). Ce n'est pas toujours possible, mais c'est l'idéal.
3. Une session de formation est organisée avec les enquêteurs. Durant cette formation, on vérifie la qualité de la traduction et on s'assure que tout le monde comprend les questions de la même manière. C'est un processus itératif.
4. Toujours lors de la formation des enquêteurs, un test est conduit sur le terrain au sein des communautés parlant la langue des outils d'enquête (cela peut impliquer certaines complications logistiques si plusieurs langues sont nécessaires).
5. Suite au test sur le terrain, toutes les réponses à l'enquête sont examinées et l'on procède aux derniers changements au niveau de la traduction. On demande également aux enquêteurs de commenter les réponses qu'ils ont obtenues, et ils ont alors une dernière possibilité de réviser la traduction et de clarifier tout malentendu.

Source : Kathy Duryee, WVI, communication personnelle, 2014

## Les enquêtes en ligne

Les enquêtes en ligne sont assez faciles à mettre en place et réaliser. Elles ont largement remplacé les enquêtes par courrier électronique et certaines formes d'enquête téléphonique.

### Éléments clés :

- Très utilisées dans l'EAH.
- Un outil très utile pour connaître les points de vue de vastes groupes. Mieux adapté à la collecte des points de vue du personnel des organisations chargées de l'exécution ou partenaires, mais peut également collecter les points de vue de la communauté humanitaire au sens large.
- Il est possible de répondre à l'enquête à distance quel que soit le fuseau horaire.
- Généralement actives pendant deux à quatre semaines.
- Peuvent rassembler des données qualitatives ainsi que des données quantitatives limitées.
- Les personnes interrogées sont libres de participer ou non si bien que les taux de réponses peuvent être assez faibles. Les personnes interrogées sont généralement ciblées en tant que groupe plutôt que sélectionnées de façon

aléatoire, ce qui signifie que les résultats ne doivent pas être considérés comme représentatifs de la population au sens large.

- Convient pour identifier de potentiels informateurs clés – demandez toujours aux personnes interrogées si elles souhaitent être recontactées pour fournir de nouvelles informations.
- Les personnes interrogées peuvent être invitées, mais elles choisissent elles-mêmes de participer ou non en général, si bien que les résultats des enquêtes en ligne ne doivent pas être généralisés.
- Peu onéreuses.
- Le nombre de réponses va en général de quelques dizaines à plusieurs centaines, voire plus.
- Une question éthique est posée par le fait qu'une pratique commerciale courante consiste à proposer aux participants de l'enquête un avantage, comme la possibilité de gagner un bon d'achat ou un appareil électronique, mais cela n'a pas cours dans l'EAH.



#### Astuces :

- Limitez le nombre de questions pour augmenter le taux de réponse.
- Au début de l'enquête, précisez le nombre de questions ou le temps que devrait prendre le fait d'y répondre.
- Envoyez plusieurs rappels pour augmenter le taux de réponse. Si vous envoyez des invitations individualisées, le site Internet de l'enquête peut savoir qui a déjà répondu et ne pas leur envoyer de rappel.
- Réaliser une enquête en ligne durant la phase de démarrage vous permet d'identifier de potentiels informateurs clés ayant des points de vue tranchés ou des analyses bien argumentées.
- Traduire l'enquête dans les langues appropriées peut augmenter la participation du personnel national.
- Dillman *et al.* (2009) est l'ouvrage de référence en matière d'enquêtes en ligne.

## Exemples d'utilisation d'enquête en ligne

- L'étude Stay and Deliver portant sur les bonnes pratiques en matière de mise en œuvre de l'assistance humanitaire dans différents contextes de sécurité et de risques a utilisé une enquête en ligne multilingue (anglais, français, espagnol et arabe) pour collecter les opinions de 1.148 acteurs locaux (Egeland *et al.*, 2011).
- L'évaluation de la réponse du HCR face à l'arrivée de réfugiés au Liban et en Jordanie a utilisé une enquête en ligne pour valider les résultats de l'évaluation (Hidalgo *et al.*, 2015).
- L'évaluation du projet ECB (Ky Luu *et al.*, 2014)



#### **Astuces :**

Vérifiez si une organisation a réalisé une enquête au moment du départ de son personnel – cela peut être utile pour identifier les sujets clés d'une évaluation. Toutefois, la confidentialité doit être respectée.

La FICR s'est servie de telles enquêtes et en a mis en place pour des opérations d'urgence à grande échelle.

Nous avons fait cela pour Haïti et les Philippines. Dans les premières semaines qui suivent la catastrophe, nous posons un nombre limité de questions à notre personnel qui quitte l'intervention : leur perception, dans quelle mesure nos principes fondamentaux sont respectés ou non, dans quelle mesure nous prenons en compte les points de vue et opinions de la population affectée. On laisse le processus se poursuivre pendant plusieurs mois. Ainsi, lorsque l'on réalise une ETR ou une évaluation humanitaire, nous sommes capables de fournir aux évaluateurs de nombreux éléments sur notre personnel toujours sur le terrain ou ayant participé à l'intervention. Il est utile de rendre cette enquête anonyme et c'est vraiment une pratique que j'ai trouvée pertinente. (Josse Gillijns, communication personnelle, 2015)

## 13.5 L'observation

L'observation est une technique utile pour l'évaluation et elle se situe au cœur de la méthode de l'EAH parce qu'elle peut aider l'évaluateur à mieux comprendre le contexte. L'observation peut être structurée ou non structurée.

### **L'observation structurée**

L'observation structurée consiste à ce que l'observateur utilise un formulaire pour enregistrer ce qu'il remarque. Elle a peu été utilisée dans l'EAH mais il pourrait l'être plus.

Par exemple, si une organisation a formé des familles à la collecte d'eau potable, on peut placer à la pompe un observateur chargé d'enregistrer les comportements afin de juger l'efficacité de la formation. L'observateur aurait un formulaire composé d'une première colonne comportant les comportements suivants et d'une colonne pour chacune des personnes utilisant le puits. Si le comportement promu est observé, la colonne de l'utilisateur en question est cochée.

Numéro d'utilisateur :	1	2	3	4	5	6	7	8	9	...
Élément de bonnes pratiques										
Conteneur scellé apporté										
Conteneur scellé ouvert à l'emplacement du puits										
Capuchon maintenu à l'écart des salissures										
Conteneur rincé										
Eau de rinçage mise dans le drain d'évacuation										
Pas d'éclaboussures excessives										
Conteneur scellé avant de repartir										
Numéro de la note : (si observation spéciale)										
<b>Notes :</b>										

Alors que certains systèmes d'observation structurée peuvent être complexes,<sup>6</sup> d'autres peuvent être relativement simples. Bentley *et al.* (1994) ont écrit un excellent guide sur cette technique qui comprend notamment plusieurs formulaires d'échantillonnage et des conseils pour les adapter à chaque situation. Même si l'observation structurée est le plus souvent utilisée dans les domaines de l'assainissement et de l'éducation, elle pourrait être utilisée de nombreuses autres manières dans l'EAH, notamment :

- Les enquêtes d'utilisation des équipements, notamment les points d'eau, les latrines municipales, etc.
- Les différences de traitement entre hommes et femmes, garçons et filles aux points de service, comme les centres de distribution ou les cliniques.
- La gestion des différents plaignants et types de plainte par les bureaux des plaintes.
- Le niveau d'attention et d'interaction durant les réunions de coordination.

Toutes les observations structurées ne nécessitent pas de coder les données. Dans certains cas, l'enquêteur compte simplement le nombre de personnes en train de faire quelque chose de spécifique sur une période de temps définie. L'enquêteur n'a pas besoin de compétences particulières autres que la capacité à réaliser une tâche souvent ennuyeuse. Un formulaire d'observation structurée peut être développé pour tout objectif – par exemple pour observer les interactions aux réunions communautaires – en fonction des questions d'évaluation.

Si les participants sont conscients de la présence de l'observateur, cela peut créer de la réactivité. Au Bangladesh, Ram *et al.* (2010) ont montré que la présence d'observateurs augmentait le lavage des mains. Ils ont pu mesurer cela en installant des savonnettes équipées de détecteurs de mouvement. Le savon était bien plus manipulé les jours où le lavage des mains était observé, et 22 % des ménages ne se sont même pas déplacés les autres jours.

## L'observation non structurée



### Astuces :

Si possible, regardez autour de vous avant de vous engager dans des entretiens individuels ou de groupe avec la population affectée. Grâce à cela, il se peut que vous collectiez des informations qui vous permettent de clarifier des réponses et de promouvoir des échanges plus francs.

L'équipe d'évaluation de la Swiss Solidarity Tsunami a entrepris des visites de sites systématiques dès le début des recherches et ce, sur chaque site. Chaque membre de l'équipe a pris des notes et les a comparées à la fin de la journée avec celles des autres. Certains évaluateurs connaissaient déjà certains lieux, et cela leur a donné une bonne base avec laquelle comparer ce qu'ils rencontraient (Ferb et Fabbri, 2014: 19).

L'observation non structurée est celle où les évaluateurs prennent des notes lors de marches d'observations ou simplement en se promenant sur le site. Lee (2000) identifie cinq types d'observation simple.

- **Signes physiques extérieurs.** Comment les gens sont-ils habillés ? Y a-t-il des signes de richesse ou de détresse ou les deux ? Exemple : lorsque les chefs communautaires ont insisté durant une évaluation en Angola sur le fait que la communauté ne mangeait pas à sa faim en raison de diminutions des rations alimentaires, alors que l'on pouvait observer des sacs de graines dans de nombreux foyers.
- **Mouvement expressif.** Voit-on des enfants courir et jouer ? Exemple : une évaluation au Malawi où un village n'avait pas rencontré l'équipe de mise en œuvre de l'organisation en chantant et en dansant comme c'est la norme. Il était évident que les villageois n'étaient pas satisfaits du travail de l'équipe de mise en œuvre.
- **L'emplacement physique correspond à la manière dont les personnes organisent l'espace.** Trouve-t-on des exemples d'espace partagé ou de familles restant très à l'écart ? Les gens voyagent-ils sur les routes ou non ? Bien sûr, certains de ces éléments peuvent être déterminés culturellement mais ils peuvent donner une indication de sujets sous-jacents.
- **Conversations *in situ*.** Cela peut être de brefs échanges que les évaluateurs peuvent avoir tout en parcourant un site à pied. Les gens se plaignent-ils ? Quels sont les sujets de préoccupation dont ils font part ?
- **Comportements liés à l'emploi de son temps.** Ce point est plus facile à saisir au moyen d'une observation structurée, mais l'évaluateur peut néanmoins remarquer des choses comme les personnes qui reviennent d'une distribution alimentaire après la tombée de la nuit, ce qui pourrait laisser entendre que la distribution était mal organisée.



**Astuce :**

Emmenez un appareil photo là où cela est approprié. Les photographies sont beaucoup plus puissantes que les descriptions verbales d'observations. Prenez en photo des observations intéressantes et utilisez-les dans votre rapport.

L'utilisation de caméras pour filmer les observations et enregistrer les entretiens est également de plus en plus fréquente, voir discussion sur le consentement éclairé et la confidentialité à la [Section 14 : Impliquer la population affectée dans votre évaluation](#).





### Exemple de bonnes pratiques : Utiliser des photographies pour étayer vos messages clés

Durant la réponse à la crise du Kosovo, une organisation dotée d'une solide expérience dans l'eau et l'assainissement a réalisé un certain nombre de projets dont les performances étaient en-dessous des standards habituels. Au moment de l'introduction de la réunion de débriefing avec l'organisation, l'idée que cette dernière n'aurait pas réussi à atteindre ses propres standards a été fortement rejetée par son personnel. Toutefois, l'atmosphère a changé lorsque l'évaluateur a utilisé une série de photos - dont l'une montrait une poche d'eau de 5 tonnes en équilibre instable et située dans un environnement où évoluaient de nombreux enfants – pour illustrer la négligence envers les règles de sécurité et les politiques de l'organisation (Wiles *et al.*, 2000).

## 13.6 Les mesures discrètes

Les mesures discrètes cherchent à éviter d'interférer dans la vie de la population affectée (Webb *et al.*, 1966 ; Lee, 2000 ; Gray, 2014). Comme Patton (2014) le remarque, le but des mesures discrètes est d'éviter que la population affectée ne réagisse à la tentative de mesurer quelque chose.

Internet offre une large gamme de mesures discrètes, en particulier au niveau du suivi des visites de pages, des recherches et des réseaux sociaux. Les sujets peuvent être aussi prosaïques que le fait de rechercher comment une faute de goût vestimentaire affecte l'intérêt pour les célébrités (Pfister, 2011) que la fréquence des recherches Google pour le terme « séisme ». Bien sûr, de telles méthodes sont discrètes, mais cette section renvoie à des mesures d'observation non-réactive plutôt qu'à une analyse du contenu, que ce soit des documents conventionnels, des pages Internet ou d'autres ressources.



### Exemple de bonnes pratiques : La FICR a utilisé les réseaux sociaux comme mesure discrète lors de la réponse au cyclone tropical Haiyan qui a frappé les Philippines

Actuellement, chaque catastrophe génère un hashtag, parfois plus d'un, et en analysant le trafic sur Twitter, sur Instagram..., vous pouvez découvrir des éléments qui pourront être validés par votre personnel ou par d'autres entretiens. C'est un outil intéressant pour trianguler certains des résultats qui sortiront d'une évaluation humanitaire.

(Josse Gillijns, communication personnelle, 2015)

Webb *et al.* (1966) ont suggéré que les mesures discrètes peuvent se manifester sous la forme de : Usure, ou d' Accumulation.

L'exemple classique de mesure de l'usure donné par Webb *et al.* était l'observation de marques d'usures sur les carreaux de tapis comme indicateur de l'intérêt des visiteurs pour diverses expositions de musées. Un autre exemple de mesure de l'usure était celui où Patton a suivi la consommation de café durant une formation d'une semaine pour en déduire quelles étaient les sessions les moins intéressantes (Patton, 2014). Dans le contexte de l'EAH, l'usure de différents sentiers peut par exemple donner une idée de l'intensité de l'utilisation de différents points d'eau.

Un exemple de mesure d'accumulation était l'étude de Blake (1981) portant sur les graffitis à Hawaï comme mesure discrète des relations ethniques. Dans le contexte de l'EAH, un exemple de mesure d'accumulation pourrait être le taux de remplissage des latrines ou le taux de déchets générés.

Les mesures discrètes sont très efficaces car elles évitent le risque que l'acte qui permet de mesurer n'affecte la mesure. Alors que l'on trouve des exemples de mesures discrètes dans la gestion de l'action humanitaire, il en existe très peu dans l'EAH. Un exemple de mesure discrète de la mortalité consistait à comptabiliser le nombre de linceuls utilisés par les établissements spécialisés (pour envelopper les corps lors de la mise en terre) au sein des réfugiés installés en Tanzanie occidentale. Ces linceuls étaient librement disponibles sur demande mais les réfugiés étaient trop superstitieux pour en demander un pour toute autre chose qu'une inhumation. C'était une mesure érosive puisqu'elle réduisait les stocks de linceuls.

Un exemple de mesure discrète d'accumulation dans le cadre d'une évaluation était l'ETR inter-organisations suite aux inondations ayant frappé le Mozambique en 2007 (Cosgrave *et al.*, 2007). Dans ce cas, la mesure d'accumulation était la poussée de l'herbe dans les abris temporaires que les villageois déplacés avaient construits. Il en ressortait de façon très nette que les habitants étaient retournés dans leurs villages et ne revenaient dans le site temporaire que pour les distributions. Un exemple de mesure d'usure était celui d'un camp de réfugiés où le chemin menant aux toilettes du marché était usé alors que celui menant au point de lavage des mains situé juste à côté ne l'était pas.

L'évaluation du programme de réponse rapide aux déplacements en RDC a indiqué que le ramassage de l'herbe par les PDI pour les abris temporaires était une mesure discrète des retards pris par la livraison des bâches en plastique (Baker *et al.*, 2013: 22).

## 13.7 Les méthodes tournées vers l'apprentissage

Les méthodes listées ci-dessous conviennent tout particulièrement aux évaluations tournées vers l'apprentissage. Celles-ci sont en général assez participatives et doivent impliquer les personnes qui doivent ou veulent être impliquées dans l'apprentissage. Les méthodes doivent être propices à la réflexion et admettre les erreurs comme les réussites. La liste suivante n'est en aucun cas exhaustive.

La culture activiste des opérations humanitaires fait que l'on parle souvent d'un manque de temps pour les exercices participatifs et réflexifs. Pourtant, seule une part réduite de l'action humanitaire est dans un rapport au temps critique au point où cela la pénaliserait de consacrer du temps à une évaluation. Bien planifiées et animées, certaines activités orientées vers l'apprentissage peuvent être réalisées en quelques heures, ou une journée, et la valeur d'une pause visant à réfléchir et apprendre collectivement justifie le temps investi.

Après avoir identifié la cible de l'apprentissage – par exemple, certains membres bien définis du personnel, des partenaires ou de la population affectée –, des processus doivent être conçus de façon à pouvoir les refléter. En voici quelques exemples :

- Créer un espace sûr propice à la réflexion et à la discussion où les participants sentent qu'ils peuvent s'exprimer librement et reconnaître leurs erreurs
- Poser des questions ouvertes simples est généralement la manière la plus efficace d'encourager la réflexion (voir « Questioning and Listening » in Chapman *et al.*, 2005)
- Utiliser des processus participatifs et créatifs
- S'assurer que les leçons apprises sont bien documentées
- Fournir une bonne animation
- Encourager les personnes qui contribuent à la réflexion et à l'apprentissage à s'approprier les résultats de l'évaluation
- Mettre en place un processus de suivi.

### Le *storytelling*

Raconter des histoires est une manière naturelle de communiquer. Se concentrer sur ce processus et l'animer peut être une façon créative de faciliter l'apprentissage, en particulier lorsque cela s'accompagne d'un processus de questionnement et de réflexion visant à approfondir l'analyse. Cette section décrit deux manières d'utiliser le *storytelling* à des fins d'apprentissage : l'utilisation des métaphores et la technique du changement le plus significatif (MSC sous son acronyme anglais).



### Exemple de bonnes pratiques : Le *storytelling* utilisé par Oxfam GB à Aceh

Juste après le Tsunami qui a frappé l'Asie, Oxfam a procédé à une revue d'apprentissage interne à Aceh. Pour faciliter l'auto-apprentissage au sein du personnel et le rendre plus participatif, une approche par *storytelling* a été utilisée. Celle-ci a été animée par le conseiller Global MEAL d'Oxfam GB qui a expliqué l'exercice de la manière suivante : « Je ne voulais pas du classique « Bonjour, je suis d'Oxfam et nous sommes venus évaluer le programme » parce qu'à partir de ce moment-là vous avez déjà introduit un biais ». On a dit au personnel de ne porter aucun vêtement frappé du logo et on les a déposés en dehors de la ville. On leur a ensuite demandé de se promener au sein de la communauté, de s'asseoir à côté de quelqu'un et de simplement lui dire 'raconte-moi ton histoire' ». « Ils ont dû rester assis et écouter avant de revenir avec ce qu'ils avaient entendu. C'était super. Le personnel a vraiment apprécié, a appris des choses qu'il ne savait pas à propos du programme : des erreurs commises, des lacunes, des personnes laissées de côté, du ciblage mal fait, etc. Ils ont eu le sentiment d'en retirer beaucoup de choses, et nous avons eu une bien meilleure image de ce qui se passait au sein de la communauté que si nous avions réalisé une évaluation plus classique où l'on dit aux personnes d'où l'on vient et pourquoi nous sommes là. »

Source : Vivien Walden, Oxfam GB, communication personnelle, mars 2015

### Utiliser des métaphores

Le meilleur moyen d'utiliser des métaphores est de réaliser un atelier, par exemple en respectant les étapes suivantes :

1. Les participants, individuellement ou en groupes, dessinent des images (par exemple, un arbre ou une rivière) pour représenter leur expérience du programme en cours d'évaluation. Les différentes parties de l'arbre ou de la rivière représentent les différentes parties du programme.
2. Les participants racontent à leurs collègues une histoire à propos de leur dessin, sans être interrompus.
3. Ceux qui écoutent posent des questions sur l'histoire de manière à en approfondir l'analyse et l'apprentissage ainsi que pour proposer leurs réflexions.
4. A la fin de ce processus, le groupe liste les principales réalisations, difficultés et leçons apprises à partir de l'histoire, cela à propos de leur expérience collective du programme.
5. Cette étape peut se focaliser sur l'avenir en posant les questions suivantes à la fin :
  - a. Comment peut-on s'appuyer sur ce qui s'est bien passé ?
  - b. Comment peut-on traiter les difficultés ?
  - c. Comment appliquerons-nous ce que nous avons appris ici ?



### Exemple de bonnes pratiques : Utiliser des métaphores et le storytelling dans une étude

L'étude qu'a réalisée Tearfund en matière de renforcement des capacités liées à la gestion des catastrophes a utilisé la métaphore de l'arbre comme base d'un *storytelling*. Lors d'un atelier de suivi et d'apprentissage, il a été demandé à chaque partenaire de Tearfund de dessiner un arbre.

- Les racines représentaient les valeurs et principes qui sous-tendaient le programme.
- Le tronc représentait les partenaires (organisations et individus).
- Les branches représentaient les activités.
- Le fruit représentait les réalisations du programme.
- Les feuilles représentaient les leçons apprises.
- Les branches cassées sur le sol représentaient les difficultés internes.
- Les nuages dans le ciel représentaient les difficultés extérieures.
- Les bourgeons représentaient les activités planifiées mais pas encore mises en œuvre.
- Le storytelling a été suivi par une étude d'apprentissage.

Malgré un certain scepticisme initial, les participants ont dit qu'ils avaient trouvé ce processus participatif à la fois visuel et créatif, et qu'il avait encouragé la réflexion et l'apprentissage, amené à la prise de conscience collective des réalisations, des difficultés et des leçons apprises, et donné un aperçu de l'ensemble du programme. Ils ont aussi reconnu la difficulté de traduire l'apprentissage en action.

### La technique du changement le plus significatif (MSC)

La technique du changement le plus important est une approche participative qui peut être utilisée avec les membres du personnel ou de la communauté locale. Lorsqu'elle est utilisée avec la population affectée, elle peut se concentrer sur les résultats ou l'impact. Elle s'appuie sur le niveau du terrain de façon ascendante en utilisant des histoires qui enregistrent ce qui s'est passé, quand, pourquoi et pour quelles raisons c'était important. Cette technique a été utilisée lors de l'évaluation de la réponse de la FICR au tremblement de terre de 2007 au Pérou ; on a alors demandé aux participants d'identifier les changements les plus importants qui s'étaient produits comme résultat de l'intervention de la Croix-Rouge péruvienne (Martinez, 2009).

Les étapes sont les suivantes :

1. Les histoires sont collectées au niveau du terrain en réponse à deux questions : durant le dernier mois [ou une autre période de temps], à votre avis, quel fut le changement le plus important qui a eu lieu dans les vies des personnes qui ont participé au projet ? Pourquoi cela a-t-il été le changement le plus important ?
2. Les histoires les plus importantes sont sélectionnées par un panel de parties prenantes ou de membres du personnel.

3. Une fois que les changements ont été enregistrés, les histoires sont lues à haute voix, souvent dans le cadre d'un atelier, de manière à ce que les participants discutent et réfléchissent sur la valeur des changements rapportés.

Vous trouverez de nombreux conseils sur cette technique dans Davies et Dart (2005).

## Les ateliers

Les ateliers organisés avec du personnel sont un élément courant de l'EAH. Ils sont souvent organisés :

- Au début du travail de terrain. Cela permet à l'équipe d'évaluation de se présenter, de présenter le rapport de démarrage et d'avoir des retours du personnel. C'est aussi une manière efficace de préparer ses entretiens ultérieurs avec des informateurs clés.
- A la fin du travail de terrain. Cela permet à l'équipe d'évaluation de débrief sur ce qu'elle a trouvé et de tester résultats et conclusions auprès des participants.

L'équipe d'évaluation peut utiliser ces deux ateliers pour collecter des données, identifier des sujets clés, et affiner les résultats (de la phase de démarrage ou du travail de terrain). Comme indiqué dans la Section 5 : *Elaborer votre évaluation*, les ateliers peuvent être utilisés pour développer ou valider la TdC, et définir ce qui correspond aux notions « adéquat, bon et excellent » afin de créer une grille d'évaluation (voir [Section 10 : Les méthodes de recherche documentaire](#)).

Le rapport de l'ETR d'Oxfam portant sur la réponse à la sécheresse au Kenya reflétait « les commentaires issus des ateliers de *débriefing* organisés avec les partenaires et le personnel » (Murphy *et al.*, 2011 : 2). C'est une approche assez courante. Les ateliers de démarrage sont moins courants que ceux de *débriefing*, mais ils ont la capacité d'augmenter l'appropriation des parties prenantes de l'évaluation. Ils sont également utiles pour organiser les entretiens et la logistique des visites de terrain.



### Astuce :

Lors de l'atelier de lancement, demandez aux participants quelles questions ils aimeraient voir traitées par l'évaluation. Il ne sera pas possible de toutes les traiter mais cette discussion peut néanmoins clarifier ce que l'évaluation va vraiment faire et augmenter l'adhésion.

Les ateliers peuvent également être utilisés pour des objectifs plus vastes. L'évaluation d'impact des transferts d'espèces dans la province ouest de Sumatra a organisé un premier atelier de lancement et un second d'apprentissage pour présenter les résultats de l'étude. L'atelier d'apprentissage a également fourni l'occasion aux parties prenantes de partager des bonnes pratiques et de développer certaines mesures concrètes permettant de plaider en faveur d'un soutien des bailleurs pour les approches de transfert d'espèces (Aspin, 2010). Durant les trois mois qu'a duré l'ETR de la réponse au séisme d'Haïti, les ONG locales se sont servies d'un atelier avec les évaluateurs pour commencer une discussion sur la manière d'améliorer la collaboration avec les acteurs internationaux (Grünewald *et al.*, 2010: 18).

### Éléments clés :

- Très utilisés dans l'EAH, le plus souvent pour un *débriefing* à la fin du travail sur le terrain.
- Organisés en général avec un groupe de personnes qui sont au fait du sujet évalué.
- Souvent utilisés pour développer ou tester la TdC ou pour valider le plan d'évaluation ou les conclusions.
- Générer des données sur la validité de ce qui est proposé.
- Peuvent être utilisés pour développer des recommandations à partir des résultats et des conclusions de l'évaluation.
- Nécessitent une salle, des sièges, des *paper-boards*, un projecteur, etc.
- En général, un repas est servi.
- Les organisations des participants de l'atelier sont généralement sélectionnées de façon raisonnée mais chaque organisation décidera de qui y participera. Le principal problème avec les ateliers, c'est que les participants potentiellement les plus utiles sont souvent trop occupés pour pouvoir les suivre.
- C'est une méthode d'un coût moyen.
- En général, on dénombre un ou deux ateliers mais il peut y en avoir plus dans certains cas spécifiques pour une petite évaluation de deux semaines.

Les ateliers peuvent être moins utiles si des groupes, présentant des intérêts très différents ou sans différences majeures en termes de statut ou de pouvoir, participent au même événement. Cela complique l'atteinte d'un consensus sur des sujets particuliers.



### Astuce :

Envisagez la possibilité d'organiser des ateliers en cascade à différents niveaux. Une série d'ateliers en cascade reprend les conclusions d'un atelier pour nourrir le sujet de discussion du suivant. Par exemple, un atelier initial d'une journée destinée au personnel sur le terrain ou aux organisations partenaires peut identifier des sujets demandant une résolution à leur niveau. Le lendemain, un atelier destiné aux cadres intermédiaires peut prendre en compte ces sujets et identifier certains des problèmes sous-jacents qui ont causé ces problèmes. Il est ensuite possible de suggérer différentes manières de traiter ces sujets. Enfin, un atelier d'une demi-journée destiné aux cadres supérieurs peut examiner les résultats des deux premiers ateliers et décider quelle action entreprendre pour les résoudre.

Voir [Section 13 : Les méthodes de terrain](#) pour un exemple d'utilisation des ateliers en vue de développer des recommandations.

### Les outils d'apprentissage propres aux ateliers

Les ateliers peuvent servir à recueillir les connaissances des participants mais aussi à faire remonter à la surface les connaissances existantes des parties prenantes pour leur donner l'opportunité de les articuler et de les analyser. Un format bien connu des ateliers d'apprentissage est la revue après action, étudiée ci-dessous, mais de nombreux autres outils peuvent être utilisés lors d'ateliers pour promouvoir l'apprentissage, notamment :

- **Dotmocracy** : lorsque les participants indiquent leur approbation concernant différentes affirmations en leur attribuant des points ([www.dotmocracy.org](http://www.dotmocracy.org)).
- **Feuilles de notation d'idées** (*Idea rating sheets*) : lorsqu'un participant peut générer une idée et que les autres participants peuvent l'approuver ou non ([www.ideaatingsheets.org](http://www.ideaatingsheets.org)).  
Les feuilles de notation d'idées résolvent certains des problèmes propres à dotmocracy.
- **Mini-grilles d'évaluation** : lorsque les participants peuvent utiliser différents points de couleur pour indiquer leurs opinions sur la valeur ou d'autres aspects de composantes d'une intervention. Elles peuvent également être utilisées pour présenter rapidement les résultats de l'évaluation (Davison, 2014).
- **Tri de cartes hiérarchiques** : une solution reposant sur un classement participatif de cartes conçue pour fournir un éclairage sur la manière dont les personnes classent et notent différents phénomènes. Semblable à la notation de l'évaluation participative rapide (EPR) mais s'utilise avec des groupes alphabétisés.<sup>7</sup>
- **Images-contexte** : explorer, reconnaître et définir une situation au moyen de schémas de manière à créer un modèle mental préliminaire.<sup>8</sup>
- **La création de diagrammes** : demander aux parties prenantes de représenter les relations sous la forme d'un diagramme ou de développer des arbres à problèmes ou encore des modélisations.



Il existe bien d'autres outils, qui ont été développés pour l'EPR ou les formations, qui peuvent être utilisés pour collecter des données avec des groupes de participants spécifiques.

### Les ateliers des revues après action

Les revues après action (RAA) peuvent être utilisées comme approche alternative d'apprentissage pour l'évaluation ainsi que comme méthode pour répondre aux questions d'évaluation.



#### Définition : La revue après action

Une revue après action est un échange structuré à propos d'une intervention humanitaire qui permet à une équipe de réfléchir à ce qui s'est passé, à pourquoi cela s'est passé et à la manière de conserver les points forts et d'améliorer les faiblesses.

Une RAA est généralement un processus qui demande une animation dans le cadre d'un atelier et implique toutes les personnes identifiées comme la cible de l'apprentissage. Une atmosphère ouverte qui renforce la confiance entre les participants est essentielle pour réussir sa revue après action. Le principe général est le suivant : « pas d'attribution, pas de rétribution ». Un animateur neutre et objectif est essentiel pour assurer que la discussion reste focalisée sur les sujets à traiter, demeure positive et ne dégénère pas en autojustification ou en accusation.

La revue doit traiter les questions suivantes :

- Qu'est-ce qui devait se passer ?
- Qu'est-ce qui s'est réellement passé ?
- Qu'est-ce qui s'est bien passé et pourquoi ?
- Qu'est-ce qui aurait pu mieux se passer et pourquoi ?

Une RAA comprend les étapes suivantes. A chaque étape, des participants peuvent écrire leurs réponses sur des post-it, l'animateur les regroupant par thèmes pour réfléchir sur certains points spécifiques, ou alors le groupe peut répondre à chaque question de façon collective, l'animateur écrivant les réponses sur un *paper-board*.

- Les participants, seuls ou en binômes, mettent par écrit ce qu'ils comprennent de l'objectif ou de l'intention de l'action. (Parce que l'action humanitaire a souvent lieu dans des environnements chaotiques et que les plans peuvent rapidement devenir obsolètes, cette étape est une manière utile de découvrir si l'objectif est clair et partagé, et s'il est nécessaire de le redéfinir).
- Les participants écrivent ensuite ce qui s'est réellement passé, en travaillant éventuellement en un seul groupe pour construire une chronologie des événements clés et des changements au fil du temps, soit pour la situation,

soit pour le programme. Dans le cadre de cet exercice, il est utile de se servir de feuilles qui peuvent être placées sur un tableau ou sur un mur pour montrer la frise chronologique.

- Le groupe répond ensuite à deux questions : Qu'est-ce qui s'est bien passé ? Qu'est-ce qui aurait pu mieux se passer ?
- Cette étape est suivie par une discussion sur ce qui pourrait être fait différemment la prochaine fois, discussion à la fin de laquelle l'animateur résume tous les enseignements qui ont émergé et demande aux participants de voter pour ce qu'ils considèrent comme les trois plus importantes leçons tirées.

Les leçons apprises clés et toutes les recommandations exploitables sont documentées et communiquées à tous les participants. Il est possible de se mettre d'accord sur un calendrier, de 6 à 12 mois par exemple, pour évaluer la progression de la mise en œuvre des recommandations.

#### **Éléments clés :**

- Dans l'EAH, la RAA est généralement utilisée lorsqu'une organisation veut tirer des leçons de ses opérations.
- Ne dure généralement pas plus d'une demi-journée ou d'un jour.
- Nécessite un animateur et un preneur de notes distinct.
- Doit respecter les principes de non attribution et de non rétribution – l'organisation doit avoir une culture qui permettent l'identification de problèmes sans chercher par la suite de boucs émissaires.
- Efficace pour aider l'équipe d'évaluation à identifier des leçons clés de l'intervention.
- Nécessite une salle appropriée et le personnel clé doit faire l'effort de prendre sur son temps pour participer.
- Les participants sont sélectionnés de façon raisonnée.
- Assez peu onéreuse.
- En général, une seule RAA est réalisée par site. De ce fait, une évaluation multi-site peut comprendre plusieurs RAA.

Le guide d'USAID (2006) fournit de nombreux détails sur le processus.



**Exemple de bonnes pratiques : La revue après action inter-organisations de la réponse au tremblement de terre et aux tsunamis de 2004 dans l'océan Indien**

Cette revue (Baker, 2005) réalisée par le personnel de cinq ONG dans quatre pays affectés par le tsunami a duré deux jours et a identifié trois leçons principales :

- Le besoin d'une analyse socio-économique précoce pour aider à la programmation et au suivi du programme, ainsi que de diagnostics conjoints rapides.
- Un rôle central donné à la consultation et à la participation des communautés.
- L'importance de la planification de la préparation, notamment le besoin de renforcer les capacités locales pour les réponses d'urgence.

# 14 / Impliquer la population affectée dans votre évaluation

Les organisations humanitaires accordent une attention renouvelée au renforcement de leur redevabilité vis-à-vis des populations affectées. C'est un engagement clé du Transformative Agenda des Nations unies pour améliorer l'efficacité de l'action humanitaire. On constate un intérêt croissant pour le renforcement de la « communication avec les communautés » (CaC).<sup>9</sup> L'édition 2015 de *L'Etat du système humanitaire notait* : [...] une plus grande prise de conscience, sur le terrain, de l'importance de la communication avec les populations affectées... afin que les populations touchées par les conflits et les catastrophes ne soient pas considérées comme « purement bénéficiaires » et que les interventions soient conçues de manière à être plus centrées sur leurs besoins et leurs préférences... [mais] les progrès accomplis en matière de redevabilité envers les bénéficiaires de l'aide ne relèvent principalement que du discours plutôt que de la réalité. (ALNAP, 2015: 72 ; voir également Ashdown, 2011 ; et Brown et Donini, 2014).

Qu'est-ce que cela signifie pour l'EAH ? Comment les évaluations peuvent-elles impliquer plus efficacement la population affectée et à des phases plus précoces de la mise en œuvre ?

L'implication de la population affectée est un point faible de l'EAH malgré divers efforts pour augmenter la redevabilité vis-à-vis de ce groupe essentiel de parties prenantes – après tout, ce sont les personnes au nom de qui l'action humanitaire est entreprise.<sup>10</sup> Si nous écoutons attentivement la population affectée, il est essentiel de l'impliquer durant une évaluation.

## 14.1 Pourquoi doit-on impliquer la population affectée ?

C'est seulement en impliquant la population affectée et en écoutant leurs points de vue que nous pouvons savoir si les programmes et les projets humanitaires ont bien été pertinents par rapport à leurs besoins, et être ainsi en mesure de remplir le critère particulièrement difficile de la pertinence défini par l'OCDE-CAD. Les personnes affectées par une crise sont en général les meilleurs juges de l'efficacité du travail humanitaire. Nous ne pouvons pas améliorer la qualité de l'action humanitaire sans écouter leurs opinions. Cela signifie mettre en place une consultation de qualité au sein des évaluations.

Le Humanitarian Accountability Partnership (HAP), qui porte désormais le nom de CHS Alliance, a travaillé sur cette question de l'implication des populations affectées pendant de nombreuses années, développant des normes qui soulignent combien l'écoute des commentaires des populations affectées par une crise et la réponse qui y est faite doivent intervenir à toutes les phases du cycle de projet, de la planification et la mise en œuvre jusqu'au suivi et à l'évaluation des programmes humanitaires (HAP, 2013). Une grande part des leçons apprises fait désormais partie de la Norme humanitaire fondamentale (ou Core Humanitarian Standard sous son nom anglais) (2015).

Dans la pratique, les organisations ont différentes raisons d'impliquer la population affectée dans leurs programmes humanitaires, même si cela peut ne pas être toujours explicite. Brown et Donini (2014: 20-21) identifient trois raisons pour lesquelles les organisations impliquent la population affectée :

1. **Une raison fondée sur des valeurs ou des normes** : les organisations pensent que c'est la bonne chose à faire, par exemple pour accomplir un devoir moral, ou respecter les droits fondamentaux et la dignité des groupes affectés.
2. **Une raison instrumentale** : parce qu'elle rend les programmes humanitaires plus efficaces, en recueillant par exemple des informations pour alimenter les décisions du programme, et en répondant mieux aux besoins des personnes affectées par la crise.
3. **Une raison émancipatrice** : parce que cela répond à des inégalités structurelles et aux causes profondes des crises, par exemple en donnant voix au chapitre et capacité d'agir aux groupes marginalisés, ou de façon plus ambitieuse, en transformant les structures et dynamiques de pouvoir.

L'évaluation joue un rôle vital pour examiner l'efficacité des organisations en termes d'atteinte de ces objectifs d'implication. Lorsque la justification est implicite, une évaluation peut avoir à l'expliciter, par exemple en réalisant des entretiens durant la phase de démarrage. Comme les organisations cherchent à être *plus redevables* vis-à-vis de la population affectée, par exemple dans le cadre du Transformative Agenda, l'évaluation est vitale pour examiner si elles ont atteint cet objectif du point de vue de la population affectée. L'exemple de bonnes pratiques, p. 270 résume les résultats de la recherche conduite aux Philippines pour entendre le point de vue des communautés affectées par le typhon Haiyan, puisque de nombreuses organisations humanitaires ont investi du temps et des ressources dans le fait d'être plus redevables envers la population affectée.



**Exemple de bonnes pratiques : Ecouter la population affectée pour connaître son expérience relative aux efforts des organisations pour être plus redevables envers elle**

Grâce à la recherche réalisée entre novembre 2014 et février 2015 dans le cadre du projet Pamati Kita Common Services, les communautés affectées par le typhon Haiyan ont été consultées pour connaître leurs points de vue et expériences par rapport aux efforts des organisations humanitaires pour être redevables envers elles. La recherche a adopté une approche qualitative dans l'esprit d'un projet d'écoute. Les consultations ont révélé que la population affectée préférerait largement la communication en face-à-face plutôt que des moyens de communication plus technologiques, comme les assistances téléphoniques par SMS, qui avaient été privilégiés par de nombreuses organisations. Globalement, les populations affectées ont décrit leur relation avec les organisations internationales comme assez distantes, contrairement au point de vue des organisations qui pensaient avoir été bien plus accessibles.

Sources : Ong *et al.* (2015) ; Buchanan-Smith *et al.* (2015)

Les évaluations qui se sont impliquées de façon active avec la population affectée sont en général beaucoup plus riches que celles qui ne l'ont pas fait. Voir, par exemple, l'évaluation conjointe réalisée par CARE et Save the Children suite au tremblement de terre en Haïti en 2010, dans l'encadré d'exemple de bonnes pratiques ci-dessous. Seuls certains types d'évaluation peuvent avoir de bonnes raisons de ne pas impliquer la population affectée, par exemple une évaluation se concentrant entièrement sur les processus institutionnels plutôt que sur l'efficacité d'un programme en termes de réduction des souffrances.



**Exemple de bonnes pratiques : La méthode People First Impact (« Les populations d'abord »)**

Neuf mois après le séisme en Haïti, CARE International et Save the Children (SCF) ont commandité une évaluation conjointe indépendante de leur réponse humanitaire. Alors que l'évaluation a utilisé les critères d'évaluation de l'OCDE-CAD et des sujets transversaux pour examiner les efforts de l'aide à ce jour, elle a également fourni un aperçu de la manière dont les différents groupes représentatifs de la société haïtienne ont perçu et vécu les domaines sectoriels de la réponse humanitaire globale auxquels les deux organisations ont participé. →



Pour cela, les évaluateurs ont utilisé la méthode d'impact People First qui place l'expérience de la population haïtienne en son centre, et ils ont travaillé à partir de cette expérience pour déterminer l'efficacité des efforts des organisations dans leur ensemble, notamment ceux de CARE et SCF. Le personnel national des deux organisations a été formé pendant deux jours à des compétences portant sur la communication, l'écoute et l'animation, puis à conduire des discussions en *focus group* utilisant des questions thématiques ouvertes. Cela signifiait que la discussion pouvait être non-prescriptive et dépasser le cadre du travail et des projets des deux organisations de manière à augmenter le niveau de connaissances de la « vraie vie » des populations et aider ainsi à répondre aux questions des deux organisations : « Est-ce que nous faisons ce qu'il faut » et « Est-ce que nous faisons les choses bien ? »

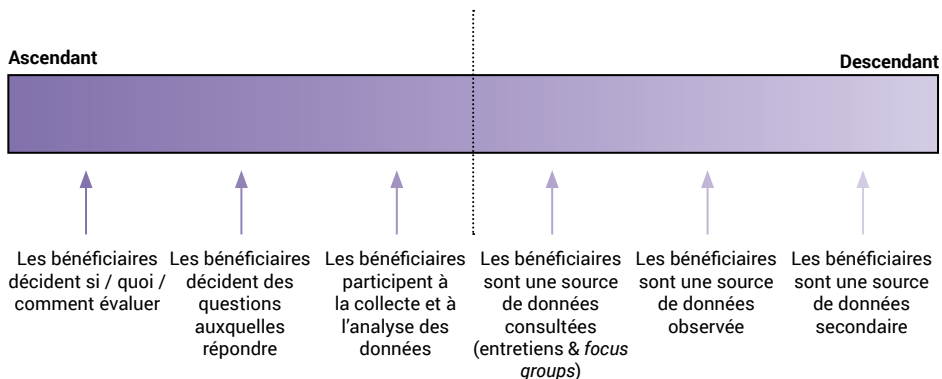
Source : O'Hagan *et al.* (2010)

Les évaluations de l'action humanitaire qui sont réellement ascendantes et participatives (Voir Section 4 : Les types d'évaluation) – lesquelles se trouvent sur la gauche du spectre du Schéma 14.1 – sont très rares dans le secteur humanitaire. Cela signifierait impliquer la population affectée dans tous les aspects de l'évaluation, depuis les phases de planification et de conception jusqu'à la collecte et l'analyse des données, l'identification des résultats et la formulation des recommandations, et enfin la diffusion.<sup>11</sup> Les sections suivantes fournissent des conseils sur les manières les plus courantes d'impliquer la population affectée dans l'EAH.

## 14.2 Planifier l'implication avec la population affectée

La population affectée n'est que rarement impliquée dès le début d'une évaluation pour définir le niveau et le périmètre potentiel de son implication même si des précédents existent. Parmi les exemples, on peut citer l'étude de faisabilité réalisée par OCHA pour une évaluation d'impact conjointe (Beck, 2011). Cela demande d'allouer du temps à la consultation *avant* que l'évaluation ne commence.

Schéma 14.1 : Le continuum participatif de l'évaluation – un exemple de la FICR



Source : IFRC-PED, 2014

Le plus souvent, impliquer la population affectée signifie planifier la consultation des bénéficiaires durant le processus d'évaluation. Comme démontré dans le continuum participatif de l'évaluation développé par la FICR, cela tend à se situer dans la moitié descendante du spectre participatif, mais c'est néanmoins important pour les raisons évoquées plus haut. Les commanditaires de l'évaluation doivent planifier ce type de consultation dès le départ en s'assurant que les ressources sont disponibles et que le calendrier permet d'importantes périodes de travail sur le terrain.

Jusqu'à présent, la consultation de la population affectée a généralement été insuffisante dans l'EAH parce que trop peu de temps et de ressources sont alloués au travail sur le terrain ; c'est en effet souvent ce qui est enlevé lorsque les calendriers doivent être raccourcis et que les budgets sont serrés. Pour autant, comme l'a souligné une étude pionnière d'ALNAP et du Groupe URD sur la participation des populations affectées par une crise dans l'action humanitaire (ALNAP, 2003), cela relève autant d'un état d'esprit que d'une série d'outils, dans ce cas un état d'esprit qui place l'implication et la consultation de la population affectée au centre du processus d'évaluation, avec les ressources qui correspondent.

Les responsables d'évaluation doivent également s'assurer de recruter des évaluateurs disposant de la bonne gamme de compétences et de connaissances. Ils doivent avoir une expérience dans la consultation des populations affectées et certains membres de l'équipe doivent avoir une connaissance préalable du contexte, et comprendre les structures communautaires et les dynamiques de pouvoir. Cela peut être particulièrement important pour une crise liée à un conflit. Si des questions sensibles de protection sont à prendre en compte, il est important que les membres de l'équipe d'évaluation aient les compétences et l'expérience appropriée pour savoir comment les gérer.



**Astuce :**

Au moment de recruter les évaluateurs qui vont mener les consultations auprès des populations affectées, en plus de les interviewer et d'examiner leur CV, cherchez à savoir s'ils ont déjà réalisé des travaux similaires, et si c'est approprié, à noter leurs références.

**Astuce :**

Au moment de planifier la consultation de la population affectée dans des contextes complexes, par exemple en raison de l'inaccessibilité et/ou de l'insécurité, prévoyez du temps pour des urgences dans le calendrier de l'évaluation au cas où le voyage serait retardé ou que d'autres arrangements ne doivent être pris.

## 14.3 Concevoir l'évaluation pour impliquer la population affectée

La première étape de la conception de l'évaluation consiste à identifier quelles questions d'évaluation des **Termes de référence** sont pertinentes pour consulter la population affectée. Les questions portant sur l'impact global de l'action humanitaire ainsi que sur son caractère approprié et pertinent par rapport aux besoins de la population affectée doivent de toute évidence être posées à la population affectée. En revanche, les questions portant sur le rapport coût-efficacité des interventions, ou sur les dispositifs internes de gestion d'une organisation, peuvent ne pas être pertinentes avec la population affectée.

Il est ensuite important d'examiner (1) ce qui existe déjà au niveau des retours exprimés par la population affectée, en particulier dans les cas où son niveau d'implication a été élevé tout au long de la mise en œuvre, par exemple grâce à l'utilisation de mécanismes de retours comme les assistances téléphoniques, le suivi post-distribution et la consultation régulière de la communauté, ainsi que (2) les lieux où ces retours ont été bien documentés. Par exemple, après le passage du typhon Haiyan aux Philippines, certaines organisations comme World Vision International (WVI) disposaient de riches bases de données informatisées qui enregistraient et analysaient les retours, ayant investi dans différents mécanismes de redevabilité.<sup>12</sup>

Il est également important de réfléchir aux groupes qui doivent être consultés au sein de la population affectée, et par conséquent à la manière dont la population doit être désagrégée. Pensez à examiner le guide UNEG *Intégrer les droits de l'homme et l'égalité des sexes aux évaluations* (2014). Par ailleurs, un bon exemple est l'évaluation UNHCR de la transversalisation des thématiques âge, genre et diversité en Colombie (Mendoza et Thomas, 2009). La désagrégation vise en général à faire ressortir différents vécus de la crise et de la réponse.

Ainsi, à titre d'exemple, les hommes et les femmes, les personnes âgées et les jeunes, les riches et les pauvres peuvent avoir vécu différemment la crise et la réponse. Observez également les fractures sociales, en particulier dans une crise liée à un conflit : par exemple, l'appartenance ethnique peut constituer une ligne de fracture majeure dans un conflit où différents groupes ethniques vivent la crise de manière très différente.

Enfin, il est nécessaire de choisir des méthodes et une approche d'échantillonnage qui permettent d'impliquer et de consulter la population. Dans la pratique, les méthodes qualitatives sont généralement utilisées pour impliquer la population affectée. Il arrive parfois que celle-ci soit consultée en utilisant des « méthodes combinées » qui mêlent enquêtes quantitatives formelles et méthodes qualitatives, ce qui peut être très efficace si c'est bien fait, mais demande beaucoup de ressources (Gayfer *et al.*, 2014). Cette méthode est plus courante dans le cadre d'évaluations conjointes dotées d'un important financement, et l'est de plus en plus dans le cadre des évaluations d'impact. Comme on l'a décrit précédemment dans la Section 12 : L'échantillonnage, l'échantillonnage raisonné est souvent utilisé à cause des contraintes liées à l'accès restreint et aux difficultés de se rendre dans certaines zones géographiques ou jusqu'à certains groupes, ce qui rend l'échantillonnage aléatoire très difficile. Dans ce cas, la connaissance de la région est primordiale pour se mettre d'accord sur l'échantillon raisonné.

## 14.4 Les méthodes d'implication de la population affectée

Le [Tableau 14.1](#) résume les cinq méthodes les plus courantes permettant de consulter la population affectée, avec leurs avantages et leurs inconvénients. L'évaluation rapide participative (EPR), également très courante, est traitée dans la sous-section suivante.

Les *focus groups* et les entretiens de groupe sont particulièrement efficaces pour écouter ce qu'a à dire la population affectée dans le cadre de discussions plus ouvertes. Ils sont en général guidés par une liste de vérification des sujets à traiter, tout en laissant de la place pour approfondir des informations nouvelles ou inattendues survenant durant la discussion. Les *focus groups* sont également une opportunité pour consulter séparément différents groupes de population, par exemple pour parler avec seulement des femmes ou seulement des hommes, ou encore avec des groupes d'âge différents. Un *focus group* doit comprendre de six à huit personnes issues d'un groupe homogène au sein de la population. Dans la pratique, il est rarement possible de maîtriser la taille et la composition du groupe, et il peut alors se transformer en un « entretien de groupe » beaucoup plus grand.

Il peut être plus approprié pour des raisons de courtoisie et de protocole de débiter la consultation lors d'une réunion communautaire visant à expliquer l'évaluation. C'est également une opportunité pour demander à la communauté son opinion générale sur le programme ou projet, mais il est peu probable que la voix de chacun soit entendue. Ainsi, les femmes, les jeunes et les groupes marginalisés peuvent être réticents à s'y exprimer.

Tableau 14.1 : Les méthodes les plus courantes pour consulter la population affectée

Méthode	Avantages	Inconvénients
<i>Focus groups</i> et entretiens de groupes	Efficace pour consulter un groupe particulier Se prête à une discussion ouverte	Difficile de contrôler la taille et la composition du groupe Requiert un facilitateur qualifié
Réunions communautaires	Utilise les structures existantes Peut impliquer un grand nombre de personnes	Peut être dominée par certains individus ou groupes Difficile de désagréger les réponses
Entretiens individuels et entretiens des ménages	Plus facile de poser des questions sensibles Peut produire des informations plus approfondies Peut être utilisée comme illustration	Plus chronophage Ne peut pas permettre des généralisations à partir d'un échantillon réduit
Entretiens semi-structurés avec des informateurs clés	Peut apporter une bonne vue d'ensemble et des éclairages susceptibles d'être approfondis lors d'entretiens individuels ou en groupe.	Implique une bonne connaissance des intermédiaires clés et des biais potentiels des informateurs clés
Enquêtes formelles	Fournit des séries de données comparables et peut être plus facilement généralisable Les données quantitatives sont plus convaincantes pour les preneurs de décision	Chronophage et relativement cher Peut ne pas être faisable dans des environnements de conflit

Les discussions en *focus group* peuvent être suivies utilement par un certain nombre d'entretiens individuels ou avec des ménages pour approfondir des thèmes spécifiques et/ou des informations sensibles qu'il n'est peut-être pas possible d'aborder en groupe. Ces discussions peuvent être particulièrement utiles en tant qu'études de cas illustratives, démontrant l'effet d'un aspect particulier de la crise et/ou de la réponse au niveau des ménages, qui est ressorti régulièrement, par exemple au moyen des discussions en *focus group*.

Les entretiens avec les informateurs clés sont couramment utilisés dans l'EAH comme une manière de consulter la population affectée. Les informateurs clés sont généralement des individus qui occupent dans la communauté une position leur permettant de commenter la réponse humanitaire globale et ses conséquences, par exemple qui a reçu ou non l'aide humanitaire et pourquoi. Ils peuvent également parler de certains effets non intentionnels de la réponse, qu'ils soient positifs ou négatifs. On peut citer comme exemples d'informateurs clés les agents de santé et les enseignants qui ont de nombreux contacts avec les membres de la communauté et occupent une position intéressante pour observer les changements et les tendances mais aussi savoir quels groupes sont plus vulnérables et marginalisés. Suite au tsunami qui a frappé la Thaïlande, les évaluateurs ont remarqué que les moines et les policiers étaient des informateurs clés très instructifs, même s'ils n'ont reçu aucune aide humanitaire.

Les enquêtes formelles peuvent également être utilisées comme une forme de consultation de la population affectée même si, comme décrit dans la Section 13 : Les méthodes de terrain, elles renvoient à des questions prédéfinies et ne permettent pas de discussions ouvertes.



## L'évaluation participative rapide

### Définition : L'évaluation participative rapide

Les techniques de l'évaluation participative rapide (EPR) (également appelée « évaluation rurale participative ») sont un groupe de méthodes permettant aux populations locales de renforcer, d'analyser et de partager leurs connaissances et leur apprentissage de façon à ce que des chercheurs venus de l'extérieur puissent les comprendre facilement.

Même si les techniques EPR ont été initialement conçues pour un contexte de développement, elles conviennent tout à fait pour impliquer population affectée.

### Éléments clés :

- Les méthodes EPR génèrent des données majoritairement qualitatives.
- Utilisées dans l'EAH pour consulter la population affectée.
- Le matériel nécessaire dépend de la méthode retenue.
- Les sites sont sélectionnés de façon raisonnée.
- Les participants peuvent se choisir eux-mêmes ou être issus d'un groupe particulier.
- Peuvent avoir lieu dans le cadre d'un entretien de groupe.
- Le nombre de sessions d'EPR peut aller de quelques-unes à plusieurs dizaines. Un plus grand nombre de sessions EPR sera nécessaire si c'est la principale technique utilisée dans le cadre d'évaluations participatives.



### A garder à l'esprit

Peu onéreuses pour les évaluateurs, mais très chronophages pour les participants. Est-ce approprié ou raisonnable par rapport à la phase de la crise et/ou au contexte en cours?

Un point fort de la plupart des techniques EPR est qu'elles reposent sur la communauté, permettant ainsi aux évaluateurs de saisir les connaissances de la communauté et non de seulement un ou deux de ses membres. La plupart de ces techniques sont très visuelles et accessibles aux groupes analphabètes. La participation aux exercices EPR a toutefois un coût réel pour la communauté. Comme le remarquait l'évaluation réalisée par Swiss Solidarity dix ans après le tsunami de 2004 en Asie durant la recherche qualitative initiale : [...] des rassemblements formels et organisés, comme les discussions en *focus group* et les outils d'évaluation rurale participative (EPR), ne génèrent pas toujours des données fiables, car les communautés ont vu passer trop de chercheurs dans les années qui ont suivi le tsunami. Par conséquent, l'utilisation de ces méthodes a été limitée aux situations où l'équipe pensait qu'elles pourraient apporter une plus-value à la recherche et alors une méthode d'entretien semi-structurée « informelle » plus ouverte a été appliquée. (Ferb et Fabbri, 2014: 18-19)

Tableau 14.2 : Les techniques courantes d'EPR

Technique	Description	Utilisation potentielle
<b>Calendrier</b>	Le groupe construit un diagramme qui montre les changements au fil du temps (par exemple, dans le travail agricole, les charges de travail liées au genre ou les maladies).	Pour les femmes, un calendrier de 24 heures peut montrer combien de temps elles passent à obtenir des produits des secours (comme l'aide alimentaire) durant le premier stade d'une crise.  Un calendrier saisonnier peut montrer les périodes de plus grande pénurie alimentaire, qui peuvent aussi être comparées dans l'évaluation avec le timing de la distribution alimentaire.
<b>Chronologie</b>	Le groupe construit une chronologie des événements.	Une première chronologie peut enregistrer les événements clés ou les moments d'insécurité durant une crise liée à un conflit, et une seconde peut enregistrer les moments où l'aide humanitaire a été apportée.
<b>Empilement proportionnel</b>	On donne au groupe une pile de 100 cailloux, haricots ou graines pour représenter un total lié à une catégorie donnée (comme les revenus d'un ménage) et on lui demande de diviser l'ensemble de manière à illustrer l'importance relative de différents éléments au sein de cette catégorie (comme les sources de revenus).	Le fait de connaître l'importance relative de différentes sources de moyens d'existence peut indiquer l'importance relative d'une intervention d'urgence spécifique comme un programme de travail-contre-argent ou d'aide alimentaire.  Cette technique peut également être utilisée pour identifier les plus pauvres de la communauté.
<b>Classement</b>	On demande au groupe de classer différents sujets, soit les uns par rapport aux autres, soit de façon bidimensionnelle en fonction de certains critères.  En utilisant le classement par paires, une matrice est élaborée avec les mêmes sujets sur les deux axes et l'on demande au groupe de classer chaque sujet par rapport aux sujets qui sont sur l'autre axe.	Le classement peut être utilisé pour comprendre comment différents types d'aide humanitaire (comme l'aide alimentaire, les biens non alimentaires, les semences et les outils) satisfont les besoins des bénéficiaires. S'il est fait de façon bidimensionnelle, il peut saisir les différents besoins d'un ménage (par exemple, des personnes âgées, des femmes, des hommes ou des enfants).
<b>Marche d'observation</b>	Les évaluateurs marchent dans le village avec un petit groupe ou avec des informateurs clés et lui/leur demande(nt) ce qu'il(s) observe(nt) (par exemple, qui a accès aux pâturages pour faire brouter son bétail ou qui vit dans telle maison).	Cela peut aider des évaluateurs à comprendre les différents impacts d'une catastrophe naturelle sur différentes zones ou différents groupes d'une communauté et à approfondir la réflexion sur la sensibilité de la réponse humanitaire par rapport à ces différences. Cela fournit également de bonnes opportunités d'observation.
<b>Cartographie sociale</b>	Méthode visuelle pour montrer l'emplacement relatif des ménages et la répartition de différents types de personnes (comme les hommes, les femmes, les adultes, les enfants, les propriétaires terriens, les sans terre, les personnes alphabétisées et les analphabètes) en même temps que la structure sociale et les institutions d'une zone.	Cela peut aider des évaluateurs à comprendre comment sont liées la localisation et la vulnérabilité au sein d'une communauté. Cela peut aider des évaluateurs à se concentrer sur des groupes particuliers pour répondre à des questions de couverture.
<b>Diagramme de Venn</b>	Les cercles qui représentent différentes catégories se recoupent là où les catégories ont une valeur en commun.	Ces diagrammes sont habituellement utilisés pour illustrer différentes institutions - par exemple, les organisations humanitaires internationales, nationales et locales - et montrer la manière dont elles sont interconnectées ainsi que leur importance relative dans la réponse à une crise.

11 / Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation 192 - 213

12 / L'échantillonnage 214 - 230

13 / Les méthodes de terrain 231 - 267

14 / Impliquer la population affectée dans votre évaluation 268 - 282

15 / L'accès restreint 283 - 292

16 / L'analyse 293 - 321

Vous pouvez trouver de plus amples informations sur l'utilisation des méthodes EPR dans *Methods for Community Participation* (Kumar, 2002) et *Participatory Rapid Appraisal for Community Development* (Theis et Grady, 1991). L'approche EPR découle du travail de Robert Chambers (1994). Des ressources ont été mises à disposition en ligne par la Banque mondiale (World Bank, 2007) et l'Agence canadienne de développement international (CIDA, 2005).



#### **Astuce :**

Les méthodes EPR ont l'air simple mais elles requièrent de bonnes compétences d'animateur. Elles doivent être utilisées avec un groupe relativement homogène, comme des femmes récemment déplacées ou des habitants restés dans un camp ou une ville après qu'il/elle a été attaqué(e). Il peut être utile de réaliser des exercices EPR en séparant les hommes et les femmes, ou encore les chefs et les autres personnes.

L'animation EPR demande de trouver le juste équilibre entre émancipation de la communauté et accaparement du processus par l'élite (Adbullah *et al.*, 2012).



#### **Exemple de bonnes pratiques : L'EPR lors d'une évaluation participative en Ouganda**

L'évaluation du programme Farmer Field School (FFS) en Ouganda (Foley, 2009) a fait un usage très complet des outils EPR. Il s'agissait d'une intervention de sécurité alimentaire et d'activités liées aux moyens d'existence semblable à de nombreuses interventions de relèvement mis en œuvre dans des contextes humanitaires.

Les douze groupes du programme ont utilisé :

- Le classement des richesses pour comprendre les perceptions de la richesse et comment elles ont pu être reflétées dans le ciblage du programme.
- Le calendrier saisonnier pour voir les cycles de production et les besoins d'intrants, ainsi que pour juger de la manière dont le programme s'alignait sur ces éléments.
- L'empilement proportionnel pour approfondir la question de l'alimentation des ménages, des revenus et des dépenses, ainsi que pour suggérer quels types de changement peuvent être attendus de la part d'un programme FFS plus efficace.
- Le classement par paires pour approfondir la sévérité relative des stratégies de survie et pour relier les idées qui en ressortent à celles relatives à la vulnérabilité et à la saisonnalité.

## 14.5 Les difficultés propres à l'implication de la population affectée dans l'EAH

Les problèmes liés à l'implication de la population affectée, et les possibles solutions, sont présentés dans le [Tableau 14.3](#) ci-dessous.

**Tableau 14.3 :** Les difficultés couramment rencontrées dans l'implication de la population affectée

Difficulté	Solutions potentielles
Avoir accès à la population affectée, en particulier dans des crises liées à des conflits où règne l'insécurité	Enquêtes téléphoniques, réunions avec la population affectée dans des environnements plus sûrs, par exemple dans les villes où se tient un marché. Voir <a href="#">Section 15 : L'accès restreint sur l'accès restreint</a>
Barrières culturelles et linguistiques	S'assurer que l'équipe de travail sur le terrain a les compétences linguistiques adéquates ou dispose d'interprètes  Recruter des chercheurs/évaluateurs locaux au sein de l'équipe
Développer des méthodes robustes pour aller au-delà des données anecdotiques	Être attentif à la conception des méthodes qualitatives, par exemple en demandant à des pairs de les relire et de donner des conseils sur la robustesse des méthodes  Utiliser la <a href="#">triangulation</a>
La population affectée dispose d'un temps limité pour la consultation	Rester vigilant par rapport aux pressions de temps et concevoir les méthodes de consultation en fonction de cela, par exemple lorsque les personnes attendent une distribution des secours, ou le soir quand elles rentrent du travail
Les intermédiaires clés cherchent à influencer la liste des groupes que l'équipe d'évaluation sera en mesure de rencontrer	Rester vigilant par rapport aux dynamiques de pouvoir politiques et de pouvoir au sein des communautés et adapter l'approche des entretiens en fonction de cela. Voir l' <a href="#">Astuce</a> sur les intermédiaires clés <a href="#">ci-dessous</a>
Comment consulter une population traumatisée	S'assurer que les évaluateurs ont les compétences et l'expérience appropriées pour impliquer la population affectée avec sensibilité. Voir également le <a href="#">Tableau 14.4</a> .

**Astuce :**

Lorsque vous consultez la population affectée, soyez vigilant quant à l'influence de certains intermédiaires clés de la communauté, par exemple les chefs communautaires qui peuvent souhaiter influencer quels groupes ou individus l'équipe d'évaluation va consulter. Il est possible que vous deviez faire un effort particulier pour atteindre les groupes marginalisés, par exemple les personnes vivant en bordure du village ou au fond du camp, mais assurez-vous que vous ne les mettez pas en danger en leur parlant (voir [Tableau 14.4](#)). Recoupez vos informations auprès d'autres informateurs clés quant à la nature des différents groupes de la communauté.

**Astuce :**

Lorsque vous consultez la population affectée, envisagez la possibilité d'utiliser des points d'entrée différents de ceux utilisés par les organisations que vous évaluez en matière de diagnostics des besoins et de suivis en cours, cela afin d'éviter les biais et d'atteindre les groupes qui n'ont pas été entendus. Réalisez par exemple une marche d'observation au sein de la communauté/ du camp, et choisissez de façon aléatoire les ménages à visiter et interroger (voir [Section 11 : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation](#)).

Le programme Listening dirigé par CDA décrit comment : Bien sûr, nous avons respecté les protocoles et rencontré les chefs de village, les chefs communautaires, les représentants du gouvernement, etc., mais nous avons toujours demandé la permission de rencontrer d'autres personnes et parfois nous avons dû insister pour garder du temps en vue de rencontrer ces autres groupes (femmes, jeunes, marginalisés). Nous avons demandé au personnel des ONG locales et des organisations communautaires (ces dernières étant plus proches du terrain) à qui parler, et demandé des suggestions aux provideurs, enseignants, infirmiers, entrepreneurs locaux, conducteurs de tuk-tuk. (Voir Anderson, Brown, Jean, 2012)

## 14.6

### Les questions éthiques liées à l'implication de la population affectée

Plusieurs considérations d'ordre éthique sont à garder à l'esprit au moment de planifier une évaluation (Voir [Section 2 : Décider de réaliser une évaluation](#)). Cette sous-section développe les questions et considérations qui sont particulièrement importantes lorsque l'on implique la population affectée.<sup>13</sup> Le [Tableau 14.4](#) résume certaines difficultés majeures et les possibles réponses éthiques.



**Tableau 14.4 :** Les principaux problèmes et les réponses possibles lorsque l'on implique les populations affectées

Question	Réponse éthique
Les groupes de population encourent-ils des risques de mauvais traitements ou de violence ciblée ?	N'interrogez pas les personnes si cela n'est pas fait en toute sécurité.
Est-ce que les groupes politiquement puissants dominent le processus de consultation ?	Soyez vigilant aux dynamiques de pouvoir et cherchez comment atteindre les personnes marginalisées et moins puissantes.
Les membres de la population affectée sont-ils traumatisés par leur expérience de la crise ?	Mener vos entretiens avec sensibilité de manière à éviter de mettre les personnes interrogées en situation de détresse, par exemple en les amenant à revivre une expérience traumatique. Pour les questions précédant un fort risque de « retraumatisation », demandez-vous si cette information est vraiment nécessaire. Sur le plan éthique, il est inapproprié de poser des questions sensibles si les évaluateurs ne sont pas en mesure d'associer les personnes traumatisées à un service pertinent.
La crise a-t-elle généré une rupture de confiance et de relations au sein de la population affectée ?	Demandez aux membres de confiance de la communauté de présenter l'évaluation et le processus de consultation.  Assurez-vous de la transparence complète en expliquant l'objectif, les contraintes, l'utilisation attendue de l'évaluation, et comment les données seront utilisées et stockées.

En plus des considérations liées au « *Do no harm* » qui sont expliquées dans la Section 2 : Décider de réaliser une évaluation, la protection des données, la confidentialité et le consentement éclairé préalable sont des garanties importantes qu'il faut mettre en place avant les entretiens et la collecte des données primaires de toute évaluation, mais plus particulièrement encore dans l'EAH où des questions de protection sont à approfondir.

La **confidentialité** et ses liens avec la **protection des données** doivent faire partie du code éthique de tous les évaluateurs. Celui-ci comprend des procédures pour garantir la vie privée des personnes interrogées durant la phase de collecte des données. Les principes généraux qui sous-tendent la protection des données et la confidentialité lors d'une évaluation sont les suivants : chaque individu est le « propriétaire » de son vécu et les données attribuables ne sont mises à disposition de l'évaluateur uniquement sur la base d'une négociation (Kushner, 2005: 74). Ainsi, le nom d'une personne interrogée au sein de la population affectée ne doit pas être cité dans un rapport d'évaluation à moins qu'une autorisation explicite n'ait été accordée.

11 / Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation 192 - 213

12 / L'échantillonnage 214 - 230

13 / Les méthodes de terrain 231 - 267

14 / Impliquer la population affectée dans votre évaluation 268 - 282

15 / L'accès restreint 283 - 292

16 / L'analyse 293 - 321

Il est également important d'obtenir un **consentement informé** avant d'interroger des membres de la population affectée, en particulier si des questions de protection sont à prendre en compte. Cela nécessite que l'équipe d'évaluation :

- S'assure que tous les répondants potentiels, y compris les enfants et les jeunes, comprennent bien tout ce qu'implique leur participation.
- Encourage les questions et les clarifications à propos de l'évaluation avant de procéder aux entretiens ou à d'autres exercices de collecte des données.
- Accorde suffisamment de temps aux participants potentiels pour réfléchir à leur participation et prendre une décision.
- S'interroge sur les compétences de l'intervieweur (certaines équipes d'évaluation comprennent des membres dotés de compétences de base en accompagnement psychologique lorsque des questions de protection doivent être approfondies).
- Fournisse aux intervieweurs des informations relatives aux services disponibles pour les personnes interrogées (par exemple, les services sanitaires et sociaux). Soyez bien conscient que des considérations spécifiques s'appliquent à tous les exercices de collecte des données dont l'objectif est de couvrir des questions en lien avec les violences sexuelles.
- Rassure les participants sur le fait qu'ils peuvent quitter un entretien à tout moment.
- Réfléchisse sur la manière de garantir au mieux comment les intervieweurs respectent les procédures éthiques développées pour l'évaluation, et sont cohérents entre eux dans leur approche (UNICEF 2015: 131-132, d'après CP-MERG, 2012).

# 15 / L'accès restreint

## 15.1 Qu'entendons-nous par accès restreint ?

La plupart des méthodes de travail sur le terrain des EAH nécessitent que les évaluateurs réalisent des observations, des entretiens et des exercices avec les parties prenantes pertinentes. Pour autant, comme mentionné dans la [Section 1](#), les évaluateurs ont souvent peu ou pas accès aux personnes qui ont reçu l'aide humanitaire, ou même aux régions où elles vivent. Ceci est particulièrement vrai pour les contextes de conflit. Les risques vont de l'attaque où les organisations humanitaires (et ceux qui évaluent leur travail) sont une cible directe (par exemple si elles sont associées à des parties au conflit), au risque d'être prises accidentellement dans un conflit en cours, ou à la mise en danger de la population affectée en cherchant à obtenir d'elle des informations.

L'accès restreint<sup>14</sup> est clairement un problème pour les personnes qui mettent en œuvre les programmes humanitaires et pour les évaluateurs. Même si un niveau de risque relativement élevé peut être acceptable comme prix à payer pour fournir une aide visant à sauver des vies, il peut être bien trop élevé pour ce qui est de réaliser une évaluation. La décision relative à la prise de risques doit toujours être proportionnelle aux bénéfices probables.

Qu'entendons-nous par accès restreint ? Voici des exemples de différents scénarios et degrés d'accès restreint :

- Un contexte trop dangereux pour tout type d'évaluation formelle, comme certaines parties de la Somalie et de la Syrie.
- Même si des évaluateurs locaux peuvent bénéficier d'un meilleur accès parce qu'ils sont mieux acceptés et moins visibles que les étrangers, par exemple dans la majeure partie de l'Afghanistan et durant le conflit au Darfour, dans d'autres cas, comme les Balkans ou le nord des îles Moluques, le personnel national peut encourir de plus grands risques que les expatriés.
- Il est possible que l'ensemble de l'équipe d'évaluation atteigne certains groupes de la population affectée, par exemple les PDI dans les villes ou à proximité, mais pas dans les zones rurales, comme ce fut le cas durant le conflit au Darfour.
- L'accès des évaluateurs internationaux à un pays ou à certaines régions peut être sujet à des contraintes administratives, comme le besoin de visas, de permis de circuler, ou de certificats de non-objection.

L'accès restreint constitue également un défi pour le suivi des interventions humanitaires et peut limiter le suivi des données dans des environnements dangereux.

Cette section se concentre sur la manière dont on décide d'entreprendre une évaluation lorsque l'accès est restreint. Elle étudie différentes approches évaluatives dans de telles situations et en quelles circonstances une évaluation peut être réalisée à distance. Tout comme la gestion à distance est à la pointe de l'aide humanitaire, l'évaluation à distance est à la pointe de l'EAH à de nombreux titres. Un certain nombre d'organisations et d'évaluateurs ont expérimenté différentes approches, et quelques-unes des plus prometteuses sont présentées ici.<sup>15</sup> La Section 15.5 examine comment assurer la crédibilité de l'évaluation lorsque l'accès est restreint et qu'une partie de l'évaluation, si ce n'est sa totalité, doit être réalisée à distance.

## 15.2 L'importance d'une étude de faisabilité de l'évaluation

Lorsque l'accès est restreint, il est primordial de réaliser une étude de faisabilité ou d'évaluabilité avant de débiter une EAH. La Section 2 : Décider de réaliser une évaluation explique le concept de base d'une étude d'évaluabilité de manière à disposer d'une base claire à partir de laquelle avancer, cela afin de rendre l'évaluation aussi utile que possible. Si l'accès est restreint, l'étude doit se concentrer sur la faisabilité d'évaluer une activité ou un programme donné, avec crédibilité, et sur le fait de savoir si cela vaut la peine ou non.



### Exemple de bonnes pratiques : Unité d'évaluation de MSF

Lorsqu'elle reçoit une demande d'évaluation, l'Unité d'évaluation de MSF commence à réfléchir en détail aux Termes de référence et aux possibles options de conception. L'Unité pense aux ajustements qui seront peut-être nécessaires par rapport au « modèle-type » des conceptions de manière à ce que cela convienne au contexte particulier d'une opération. Dans les situations d'accès restreint comme les opérations transfrontalières en Syrie, la principale difficulté est d'obtenir des données primaires sur les résultats ainsi que des retours directs de la part des bénéficiaires. Par conséquent, la conception de l'évaluation se concentrera plus sur le processus et les activités que sur les résultats et elle devra plus s'appuyer sur les données secondaires et les perceptions. Ces options et limitations sont clairement partagées avec les commanditaires de l'évaluation, de façon à ce qu'ils puissent déterminer si une évaluation remplirait leurs besoins en apprentissage et en redevabilité ou si d'autres options doivent être envisagées.

Source : Sabine Kampmueller, Unité d'évaluation de MSF, communication personnelle (septembre 2015)



**Parmi les questions clés à poser pour répondre à cette question fondamentale, on peut citer :**

1. Quels sont les principaux risques que l'évaluation rencontre dans ce contexte ? En voici certains :
  - a. Les risques opérationnels comme la sécurité personnelle des évaluateurs et la perturbation potentielle du programme si des ressources sont redirigées vers l'évaluation
  - b. Les risques financiers si des coûts supplémentaires sont associés au fait de travailler dans un environnement dangereux,
  - c. Les risques de protection auxquels la population affectée peut être exposée si elle participe à l'évaluation
2. Quelles sont les implications de ces risques pour l'accès des évaluateurs aux populations affectées ?
3. Quelles données secondaires et autres sont disponibles si les évaluateurs ne peuvent pas avoir accès à la population affectée ?
4. Quelles autres options sont disponibles pour que les évaluateurs aient accès à la population affectée (voir [Section 15.3 sur les méthodes d'évaluation à distance](#))?
5. Comment cela affectera-t-il la crédibilité de l'évaluation ?
6. Quelles sont les alternatives possibles à une évaluation, par exemple un atelier d'apprentissage et de réflexion avec le personnel, un apprentissage entre pairs avec d'autres organisations, ou une activité évaluative plus limitée ?

Une étude d'évaluabilité peut être particulièrement importante pour savoir de façon systématique si une évaluation est faisable, lorsque des organisations ont l'impression qu'elles subissent la pression des bailleurs qui les obligent à des évaluations pour remplir une obligation contractuelle.



**Exemples de bonnes pratiques : Les études d'évaluabilité**

En 2009, le IASC a commandité une étude d'évaluabilité (Cosgrave, 2010) pour une proposition d'évaluation d'interventions humanitaires dans le centre et le sud de la Somalie. Tout en traitant les sujets qui font généralement partie d'une étude d'évaluabilité, comme la clarification des objectifs et la définition du périmètre de l'évaluation, la question clé était de savoir comment réaliser une évaluation lorsque même la forme la plus simple de suivi s'est avérée très difficile. Le principal élément de l'étude d'évaluabilité était un commentaire, fondé sur les entretiens des informateurs clés et une revue documentaire, portant sur une version provisoire des TdR de l'évaluation qui datait de 2008. L'évaluation a été réalisée avec succès en 2011 (Polastro *et al.*, 2011).





En 2014, Global Affairs Canada – qui relève du gouvernement canadien, a réalisé une « analyse des scénarios et des risques » avant l'évaluation prévue de son programme pays au Soudan du Sud (CPE). La question centrale était de savoir comment le conflit en cours au Soudan du Sud affecterait les phases de démarrage et de collecte des données. Une question subsidiaire concernait les nécessaires plans d'urgence. Une analyse systématique des scénarios a étudié la situation en cours et anticipé différents résultats possibles. Un registre des risques a ensuite été développé, différenciant les risques opérationnels, financiers, de qualité du rapport et de réputation. Sur la base de cette analyse, il a été décidé que l'évaluation devait avoir lieu et qu'un certain nombre de mesures devaient être mises en place pour atténuer certains risques.

## 15.3 Comment surmonter un accès restreint ?

Les organisations et les évaluateurs ont expérimenté plusieurs manières créatives de surmonter l'accès restreint aux populations affectées. Vous trouverez dans le [Tableau 15.1](#) celles qui sont le plus couramment utilisées.

Lorsque l'accès est restreint, il arrive que les évaluateurs dépendent fortement de données secondaires, même si celles-ci sont souvent insuffisantes dans de tels contextes.



### **Définition : *Crowd-sourcing* (approvisionnement par la foule)**


Le *crowd-sourcing* (approvisionnement par la foule) a recours à un grand nombre de volontaires soit pour collecter des données, soit pour analyser des données d'imagerie, généralement à l'aide de moyens indirects. Ce type de données est appelée données *crowd-sourcées*.

## Exemple de *crowd-sourcing* en matière de suivi : l'imagerie satellite analysée par des volontaires *crowd-sourcés*

En 2011, le HCR a mobilisé l'aide de volontaires pour baliser trois différents types d'abris informels afin qu'ils lui fournissent une estimation de la population de PDI dans le corridor d'Afgooye en Somalie. Les volontaires ont pris 3.909 images satellite en seulement cinq jours et ajouté plus de 250.000 balises.

Source : <http://irevolution.net/2011/11/09/crowdsourcing-unhcr-somalia-latest-results>.

Tableau 15.1 : Comment surmonter un accès restreint ?

Technique	Comment utiliser cette méthode	Ecueils potentiels
<p>Utiliser des chercheurs/évaluateurs locaux pour réaliser des entretiens avec la population affectée (Norman, 2012: 30-35)</p> <p>(voir Exemple de bonnes pratiques p. 289)</p>	<p>Prévoir une session de formation dès le départ (Sida, 2013) et un atelier d'analyse après la fin du travail de terrain.</p> <p>Si pertinent, demandez aux chercheurs/évaluateurs locaux d'enregistrer leurs entretiens à des fins de vérification.</p> <div style="border: 1px solid black; padding: 5px; margin-top: 10px;">  <p><b>Astuce :</b> Assurez-vous de prévoir assez de temps pour ces ateliers, en fonction des compétences et de l'expérience des chercheurs/évaluateurs locaux.</p> </div>	<p>Faisable seulement si les chercheurs/évaluateurs locaux ne sont pas mis en danger.</p> <p>Dans les conflits de longue date, il peut être difficile de trouver des chercheurs/évaluateurs qui ne sont pas perçus comme associés à l'une des parties.</p>
<p>Réaliser des enquêtes en ligne, par téléphone et/ou SMS (voir Exemple de bonnes pratiques p. 290 et Section 13 : Les méthodes de terrain)</p>	<p>Elles peuvent être utilisées pour des enquêtes relativement courtes et directes, par exemple pour obtenir des informations sur l'assistance reçue par les personnes : quand, quoi et en quelle quantité. Les enquêtes téléphoniques peuvent également être réalisées auprès du personnel sur le terrain.</p> <p>Les assistances téléphoniques peuvent constituer une opportunité pour que la population affectée soulève des préoccupations, en particulier si elles ont été mises en place durant la mise en œuvre du programme (Walden, 2013: 3).</p>	<p>Les enquêtes par SMS et en ligne sont sujettes à un biais d'auto-sélection et doivent être interprétées avec soin.</p> <p>Les enquêtes téléphoniques peuvent également être associées à des biais, par exemple au fait d'être uniquement accessibles aux personnes ayant un téléphone portable.</p> <p>Dans un environnement fortement politisé, la population locale peut ne pas faire confiance et/ou être réticente à l'utilisation d'une permanence téléphonique.</p>



← Tableau 15.1 : Comment surmonter un accès restreint ?

Technique	Comment utiliser cette méthode	Ecueils potentiels
<p>Interroger des membres de la population affectée dans des zones accessibles</p> <p>(Voir l'exemple de bonnes pratiques p. 290 et Section 13 : Les méthodes de terrain)</p>	<p>Chercher à savoir si les membres de la population affectée voyagent régulièrement jusqu'à des zones plus accessibles, par exemple des villes plus sûres où se tiennent des marchés, et s'organiser pour réaliser des entretiens et des discussions en <i>focus group</i> avec ces membres.</p> <p>Demander aux membres de la population affectée de voyager jusqu'à des zones accessibles à l'équipe d'évaluation, si cela est sans danger pour eux.</p>	<p>Peut générer des biais, par exemple s'il est jugé sûr pour les hommes de voyager mais pas pour les femmes.</p> <p>Difficile de trianguler les résultats si l'équipe d'évaluation ne peut rencontrer que certains groupes particuliers.</p>
<p>Observation à distance (Norman, 2012: 45-48)</p> <p>(voir Section 13 : Les méthodes de terrain)</p>	<p>L'imagerie satellite peut être utilisée pour vérifier des infrastructures (par exemple du bâti grâce à des programmes argent-contre-travail), ou étudier des schémas d'implantation.</p> <p>On peut demander aux informateurs clés et aux membres de la population affectée de prendre des vidéos et des photographies, en utilisant des caméras dotées d'un GPS.</p> <p>Si des programmes utilisent un suivi à distance en fournissant des tablettes au personnel, des smartphones ou des appareils photos, ce matériel peut être utilisé pour enregistrer des données.</p>	<p>Si cela peut convenir pour observer le paysage physique et les infrastructures, cela peut fournir peu d'informations sur leur utilisation, par exemple qui a accès à des infrastructures spécifiques comme les points d'eau.</p>
<p>Données approvisionnées par la foule (« crowd-sourced »), par exemple Twitter, Facebook (Sida, 2014)</p>	<p>Les données <i>crowd-sourcées</i> peuvent être utilisées pour observer le degré d'utilisation d'équipements spécifiques, à l'aide d'un suivi réalisé avec un téléphone portable.</p> <p>Vérifier si le suivi à distance a été réalisé à l'aide d'un réseau social comme Facebook et Twitter, et si/comment les données pourraient être utilisées dans le cadre de l'évaluation. (Voir <u>exemple de bonnes pratiques p. 257</u>)</p>	<p>Les personnes interrogées sont auto-sélectionnées, ce qui introduit un biais. Par exemple, les jeunes sont plus susceptibles d'utiliser les réseaux sociaux que les personnes âgées.</p> <p>Des parties en conflit peuvent réaliser des campagnes sur les réseaux sociaux.</p>

Si le programme a développé des systèmes de suivi à distance durant la mise en œuvre, il est possible de s'appuyer sur eux. Par exemple, si des centres d'appel ou des permanences téléphoniques ont été mis en place, ils peuvent être utilisés par l'évaluation pour collecter des données. Voir Exemple de bonnes pratiques p. 290.



### A garder à l'esprit

Comme pour toutes les évaluations, le processus de collecte des données ne doit pas mettre en danger les personnes.





### Exemple de bonnes pratiques : L'évaluation de la coopération de la FAO en Somalie (2007 à 2012)

Dans le cadre de cette évaluation, et en reconnaissant l'accès restreint que l'équipe avait dans de nombreuses zones géographiques de Somalie, une étude approfondie du programme argent-contre-travail (CFW) de la FAO a été commanditée. Le travail de terrain s'est déroulé sur un mois. Une chercheuse/évaluatrice internationale (anthropologue) a été désignée pour diriger l'étude. Même si elle n'a pas eu accès aux sites CFW pour des raisons de sécurité, elle a travaillé avec cinq chercheurs locaux choisis avec soin et était basée à Dollow, à proximité des sites CFW. Cinq jours ont été alloués à la formation des chercheurs locaux qui ont utilisés des méthodes qualitatives PRA pour la collecte des données.

Une technique dite en « boule de neige » a été appliquée (voir [Section 12 : L'échantillonnage](#)) : elle permettait de collecter de l'information de façon itérative pendant un mois. Les chercheurs locaux ont visité chaque village pendant trois jours et sont retournés à Dollow pour le débriefing et pour parler de leurs résultats avec le chef d'équipe. C'était l'occasion d'examiner des sujets inattendus qui génèrent souvent des questions complémentaires à poser durant la prochaine visite de terrain. Les techniques et outils EPR ont donné lieu à divers supports qui pouvaient être facilement analysés et se sont révélés essentiels pour la recherche itérative et le processus de formation. Les chercheurs locaux ont réalisé des triangulations sous la forme de photographies et d'enregistrements audio. En utilisant les données du centre d'appel de la FAO mis en place à des fins de suivi durant la mise en œuvre du programme, un autre consultant somalien a été employé pour réaliser des enquêtes téléphoniques qui ont également servi à trianguler les informations collectées sur le terrain avec les bénéficiaires CFW d'autres régions.

En plus de l'étude CFW, l'équipe d'évaluation a également demandé à certains « intermédiaires » (qui pouvaient faire le voyage), par exemple des membres d'associations professionnelles de paysans et d'éleveurs, de se rendre à Mogadiscio pour rencontrer l'équipe dans un espace sûr.

Sources : Buchanan-Smith *et al.* (2013) ; Tessitore (2013)



#### Astuce :

Demandez-vous si certains évaluateurs peuvent encourir moins de risques, en raison par exemple de leur nationalité ou de leur appartenance ethnique. Cependant, gardez à l'esprit que vous avez un devoir de vigilance envers tous les évaluateurs.



### **Exemple de bonnes pratiques en matière d'entretiens par téléphone : l'évaluation d'Oxfam GB de la réponse à la sécheresse en Somalie (2011-2012)**

Pour son programme de distribution d'espèces, Oxfam GB a collecté les numéros de téléphone portable de bénéficiaires participants (l'ETR suggère que 10-15 % des bénéficiaires ont communiqué leur numéro de téléphone) et durant l'ETR réalisée en septembre 2011, un petit nombre de personnes a été appelé pour faire des retours sur le programme. Ce processus a donné lieu à des résultats intéressants. Par exemple, sur 12 numéros appelés, cinq ont répondu (en fait, une personne a rappelé après avoir manqué l'appel). L'objectif de ces conversations était d'évaluer la connaissance qu'avaient les bénéficiaires de leurs critères de sélection, leur impression par rapport au processus, leur compréhension du projet, et s'ils savaient comment déposer une plainte. Les retours obtenus par ces appels ont permis d'éclairer plusieurs points. Par exemple, toutes les personnes interrogées ont dit qu'elles comprenaient les critères de sélection, la plupart ont indiqué qu'elles savaient comment contacter un représentant de l'organisation si nécessaire, mais aucune ne savait quand et comment elle recevrait de l'argent, ni combien.

Source : Featherstone (2012: 13-14)



### **Exemple de bonnes pratiques : Amener des membres de la population affectée jusqu'à des zones plus accessibles**

Pour l'équipe du Groupe URD qui réalisait une évaluation d'un programme PDI en RDC, il était important de parler directement à la population affectée. Cependant, obtenir l'accès aux camps aurait impliqué de franchir le territoire tenu par les rebelles. Alors que les hommes étrangers couraient le risque d'être kidnappés, les hommes PDI pouvaient se déplacer de façon relativement sûre à travers la zone. Auparavant, les partenaires locaux avaient déjà profité de cette situation pour distribuer des produits de l'assistance. Par conséquent, le Groupe URD a développé des critères précis pour aider la population PDI à sélectionner différents représentants (par exemple : agriculteurs, chefs religieux ou traditionnels). Quatre ou cinq représentants par camp ont ainsi été sélectionnés et priés de se rendre dans un village sûr pour être interviewés. Les PDI avaient également la possibilité de désigner une personne se trouvant déjà dans le village qui pourrait parler en leur nom de façon rigoureuse.

Pour que ce voyage soit vu comme bénéfique, les partenaires locaux ont offert aux représentants des objets qu'ils pouvaient rapporter à leurs communautés.

L'inconvénient de ce type d'approche est que les évaluateurs n'ont pas de contrôle sur les personnes que la communauté choisira. Enfin, seuls ceux qui peuvent traverser la zone dangereuse peuvent se rendre aux entretiens (par exemple, les femmes ne peuvent pas voyager dans certains cas).

Source : Bonaventure Sokpoh, Groupe URD, communication personnelle, 2015



### Exemple de bonnes pratiques : L'évaluation à distance qui s'appuie sur des entretiens du personnel (MSF au Pakistan)

En 2009, MSF a lancé une évaluation intersectionnelle de ses projets gérés à distance dans le monde entier. En raison de problèmes sécuritaires au Pakistan, l'évaluateur n'a eu qu'un faible accès au programme ou à la population affectée. Pour avoir une image plus complète du programme, l'évaluateur a contacté du personnel actuel et passé de MSF. Le personnel international était facile à retrouver (et interrogé lors de réunions ou par téléphone), mais ce n'était pas le cas pour le personnel national. L'évaluateur a eu recours à la technique dite « boule de neige » pour identifier le personnel qu'il serait utile d'interroger et pour obtenir ses coordonnées. C'était un processus chronophage mais qui a porté ses fruits car il a considérablement augmenté l'échantillon des personnes interrogées. Avant la mission sur le terrain, les évaluateurs ont assuré le suivi des pistes fournies, mais une grande part du travail a été réalisée en capitale sur place et depuis un bureau de coordination du projet (à Peshawar). Il était important d'aller à Peshawar pour garantir la plus grande proximité possible vis-à-vis de l'actuel site du projet, cela de manière à trouver les personnes pertinentes pour les entretiens et pour mieux comprendre le contexte.

Source : Mzia Turashvili, Unité d'évaluation de MSF, communication personnelle, novembre 2015

## 15.4 La crédibilité de l'évaluation à distance

Si les évaluateurs n'ont eu qu'un accès limité à la population affectée, cela pose la question de la manière dont la crédibilité des résultats sera garantie.

Tenez alors compte des éléments suivants :

1. Dans le rapport d'évaluation, explicitez les limitations et contraintes rencontrées par l'équipe d'évaluation, ainsi que la manière dont cela a affecté les résultats de l'évaluation. Soyez clair sur les groupes de population ou les zones géographiques auxquels les résultats renvoient. Dans votre analyse, faites attention à ne pas généraliser pour des groupes et des zones que l'équipe n'a pas été en mesure de visiter.
2. La triangulation est toujours importante dans l'EAH. Lorsque l'accès est restreint, la triangulation devient encore plus cruciale. Par exemple, si vous dépendez d'équipes d'évaluateurs locaux qui ont travaillé de façon autonome, comparez les données rassemblées par différents évaluateurs sur le même site pour identifier tout biais d'évaluateur. Triangulez les informations collectées auprès de la population affectée à l'aide d'enquêtes à distance, comme les enquêtes en ligne et par téléphone, avec celles issues d'informateurs clés de la zone et du personnel de l'organisation.



3. Lorsque vous trouvez que vos résultats ne sont pas concluants à cause des contraintes auxquelles vous avez fait face, pensez à convertir des résultats incertains en hypothèses à tester par l'organisation après l'évaluation. Il se peut que vous recommandiez de nouvelles recherches et collectes de données sur des sujets précis.

**Astuce :**

Si vous obtenez des informations contradictoires de différentes sources que vous ne pouvez pas réconcilier, et que vous n'avez pas été en mesure de visiter les personnes ou zones concernées, soyez prêt à ne pas utiliser ces informations dans votre analyse, ou à expliciter dans votre rapport que vous n'avez pas pu concilier deux (ou plus) interprétations différentes.

## 15.5 Les autres options à prendre en compte si une évaluation à distance est irréalisable

- **Suivi à distance** : pensez à utiliser le suivi à distance si l'évaluation à distance n'est pas jugée faisable. Le suivi à distance devrait être plus facile que l'évaluation à distance si elle se concentre principalement sur les moyens et les produits, plutôt que sur les résultats et l'impact global. Un suivi réalisé par une tierce partie est également à envisager, par exemple si une ONG demande à une autre de réaliser son suivi.
- **Apprentissage par les pairs** : les organisations qui travaillent dans une zone relativement inaccessible peuvent être invitées à se réunir pour partager leurs leçons apprises et leurs expériences. Cet atelier pourrait également être utilisé pour encourager la réflexion et l'apprentissage. En 2013 et 2014, ALNAP a co-organisé avec le DEC des ateliers d'apprentissage par les pairs pour le personnel d'organisations travaillant sur les questions d'évaluation, de S&E, de redevabilité et d'apprentissage en lien avec la réponse en Syrie. C'était une occasion utile pour demander conseil et partager des leçons apprises (Sida, 2013, 2014).

Dans certains cas, vous pourrez envisager de conseiller que les activités de suivi à distance soient améliorées avant d'aller plus loin au niveau d'une évaluation. Il existe un certain nombre de ressources et de leçons apprises en matière de suivi à distance, notamment :

- Sur la Somalie, voir Polastro *et al.* (2011 : 29),
- Sur la réponse en Syrie, voir la revue d'apprentissage 2014 d'ACF (2014: 54-55),
- Sur un projet de recherche financé par le HIF, voir le rapport Tearfund sur les pratiques de suivi et de redevabilité des projets gérés à distance qui sont mis en œuvre dans des environnements opérationnels volatiles (Norman, 2012).

# 16 / L'analyse

Cette section se concentre sur les méthodes utilisées pour transformer les données collectées en résultats qui forment la base des conclusions et des recommandations de l'évaluation.

Cette section couvre :

- L'analyse nécessaire pour répondre à des questions évaluatives.
- L'analyse nécessaire pour répondre à des questions causales.
- L'analyse des données qualitatives issues de données primaires comme les notes d'entretiens.
- L'analyse des données qualitatives issues de données secondaires comme les rapports d'étapes.
- L'analyse des données statistiques issues de données primaires comme les données d'enquêtes.
- L'analyse numérique de données secondaires, comme celles des registres de données des distributions.
- Le processus qui va des résultats aux conclusions et aux recommandations.



## **Définition : Données primaires**

Les données primaires sont les données collectées pour les besoins de l'évaluation.

Les entretiens et l'enquête réalisés spécifiquement pour l'évaluation sont des exemples de données primaires.



## **Définition : Données secondaires**

Les données secondaires sont les données collectées pour d'autres besoins mais utilisées par l'évaluation.

Les données secondaires comprennent les documents de politiques des organisations, les rapports intermédiaires, les autres évaluations et toute littérature scientifique pertinente. Voir [Section 11 : Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation](#) pour plus de détails sur la documentation secondaire.

Les données secondaires et primaires fournissent les éléments de preuve des évaluations.



## **Définition : Éléments de preuve**

Les éléments de preuve sont l'ensemble des faits ou information disponibles capables d'étayer une proposition ou une idée particulière.



Voir le document d'ALNAP sur les éléments de preuve pour plus de détails sur la qualité des éléments de preuve (Knox Clarke et Darcy, 2014).

**Astuce :**

Les tableaux de synthèse des observations facilitent l'écriture du rapport en réunissant tous les éléments de preuve relatifs à une question ou un sujet spécifique en un seul et même endroit. Notez tous les éléments de preuve et les résultats émergents que vous avez collectés grâce aux entretiens et à la revue documentaire dans un tableau de synthèse des observations. C'est un tableau où l'équipe d'évaluation enregistre des éléments d'information par rapport à une question d'évaluation ou un sujet. Cela sert de pré-codage des données et permet de bien faire apparaître quels sujets spécifiques présentent ou non des éléments de preuve forts. Lorsqu'elle travaille à la formulation de ses résultats, conclusions et recommandations, l'équipe d'évaluation peut voir quelles questions ou quels sujets requièrent une base de connaissances plus solide. Les tableaux de synthèse des observations facilitent l'écriture du rapport en réunissant tous les éléments de preuve relatifs à une question ou un sujet spécifique en un seul et même endroit.

## 16.1 Analyse quantitative ou qualitative

Les approches de base en matière d'analyse des données quantitatives et qualitatives sont différentes. En général, les données quantitatives sont analysées de façon statistique, et les données qualitatives à l'aide d'un codage (voir ci-dessous pour une définition du codage). Comme indiqué dans la [Section 11](#), ces deux catégories ne sont pas hermétiques et les chevauchements sont nombreux. Certaines conceptions sont liées à des types d'analyse particuliers.

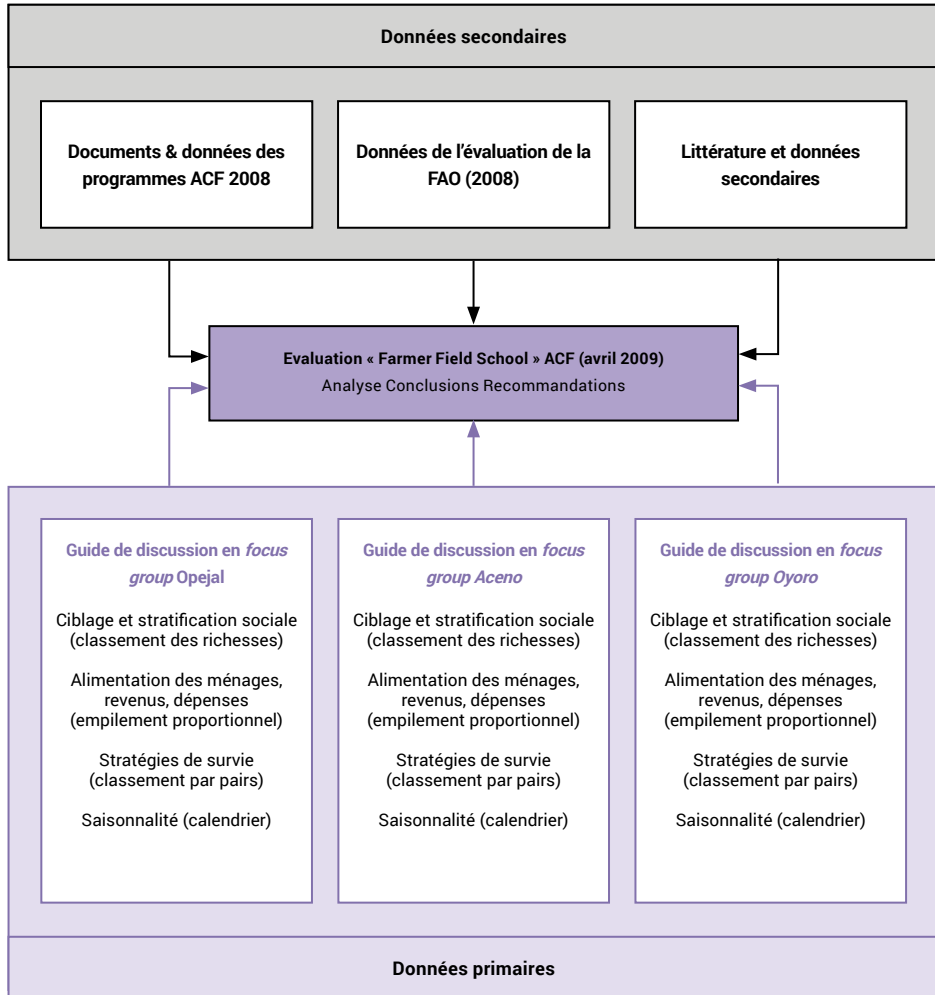
Une différence essentielle entre les méthodes qualitatives et quantitatives est le moyen utilisé pour garantir la précision et la fiabilité. Pour les méthodes quantitatives, la précision et la fiabilité sont obtenues en appliquant une méthode particulière d'une façon particulière. Pour les méthodes qualitatives, la précision et la fiabilité sont obtenues partiellement grâce à la méthode elle-même et partiellement à l'aide de la triangulation (introduite dans la [Section 13 : Les méthodes de terrain](#) sur les méthodes de terrain).

La triangulation comprend l'idée implicite que les méthodes qualitatives et combinées s'appuient l'une sur l'autre pour étayer un résultat particulier et une conclusion particulière.

Les résultats de toute évaluation doivent s'appuyer sur des sources multiples (voir [Section 10 : Les méthodes de recherche documentaire](#), et la [Section 13 : Les méthodes de terrain](#)). A leur tour, les conclusions peuvent se fonder sur de multiples résultats. La triangulation des éléments de preuve est un principe clé car elle aide à garantir leur qualité.

L'exemple qui suit, tiré d'une évaluation participative en Ouganda (Foley, 2009: 49), montre comment différentes sources d'éléments de preuve visaient à contribuer à l'évaluation.

Schéma 16.1 : Diagramme – sur la méthodologie de l'évaluation



11 / Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation 192 - 213

12 / L'échantillonnage 214 - 230

13 / Les méthodes de terrain 231 - 267

14 / Impliquer la population affectée dans votre évaluation 268 - 282

15 / L'accès restreint 283 - 292

16 / L'analyse 293 - 321

## 16.2 Analyser les éléments de preuve pour répondre à des questions normatives

Une manière de répondre aux questions normatives, comme « dans quelle mesure le programme d'abri a-t-il respecté les standards Sphère », consiste à chercher à trianguler des éléments de preuve liés à des infractions ou des conformités vis-à-vis des normes.

Alors que cette approche peut convenir pour une évaluation très basique d'une intervention simple, on peut néanmoins dire que toute conclusion négative liée à la conformité s'appuie sur des anecdotes ou des biais plutôt que sur des éléments de preuve rigoureux.

Une approche plus rigoureuse consiste à décomposer la norme pertinente applicable en éléments séparés. (Voir Section 5 : Elaborer votre évaluation pour des exemples de normes qui peuvent être applicables.) Par exemple, pour respecter les standards Sphère en matière d'abri, vous devez d'abord respecter les six standards essentiels Sphère (Projet Sphère, 2011 : 49) :<sup>16</sup>

- Intervention humanitaire centrée sur les personnes
- Coordination et collaboration
- Evaluation
- Conception et intervention proprement dite
- Performance, transparence et apprentissage
- Performance des travailleurs humanitaires

Ensuite, vous devez respecter les cinq standards Sphère minimums en matière d'abri (Projet Sphère, 2011 : 239) :

- Planification stratégique
- Planification des établissements humains
- Espaces de vie couverts
- Construction
- Impact sur l'environnement

Le fait d'examiner dans quelle mesure une intervention respecte chacune de ces normes permet de répondre à la question générale. Les standards Sphère comprennent des indicateurs qui peuvent être directement utilisés avec certains standards (si les indicateurs sont jugés appropriés dans le contexte donné). Il est possible de répondre à d'autres standards en réalisant seulement un jugement évaluatif, et les techniques décrites dans la section 16 peuvent dans ce cas être utilisées.



## 16.3 Analyser les éléments de preuve pour des questions évaluatives

Les questions évaluatives (voir Section 6 : Choisir les questions d'évaluation) impliquent que les évaluateurs forment des jugements de valeur. Elles peuvent poser un problème aux évaluateurs car de tels jugements peuvent être critiqués en raison de leur trop grande subjectivité. C'est une question qui se pose pour tous les types d'évaluation, quelle que soit la conception ou la méthode. Par exemple, une enquête peut avancer qu'un facteur critique donné a augmenté de 20 % en raison d'une intervention, mais l'évaluateur doit formuler un jugement quant au fait de savoir si cela représente une bonne ou une mauvaise performance au vu des circonstances.

L'une des façons d'aborder les questions évaluatives consiste à les décomposer en éléments distincts descriptifs, normatifs et autres à partir de l'élément purement évaluatif. Une autre approche consiste à adopter une approche rigoureuse et transparente pour répondre aux questions évaluatives.



### Définition : Raisonnement évaluatif

Le raisonnement évaluatif est le processus analytique par lequel les évaluateurs répondent aux questions évaluatives.

Le raisonnement évaluatif synthétise les informations relevant de la qualité et de la valeur en combinant :

- Les éléments de preuve liés à la performance réalisée pour un sujet donné et leur interprétation par rapport à la définition de ce qui est « bon » pour générer une notation de performance à ce sujet.
- Les notations de performance de plusieurs sujets pour fournir une conclusion globale sur la qualité générale de la performance (Davidson, 2014: 1).

Vous trouverez ci-dessous des exemples d'utilisation de raisonnement évaluatif à l'aide de tableaux évaluatifs comme le recommande Davidson (2014). Oakden affirme que les tableaux évaluatifs « offrent un processus pour expliciter les jugements d'une évaluation et sont utilisés pour juger la qualité, la valeur ou l'importance du service fourni » (2013: 5).



### Définition : Tableau évaluatif

Un tableau évaluatif est un tableau qui décrit à quoi devraient ressembler les éléments de preuve à différents niveaux de performance, pour certains critères d'intérêt ou pour l'intervention dans son ensemble.

Les tableaux évaluatifs ne sont pas notés de façon numérique comme les tableaux utilisés pour l'analyse de documents, mais à l'aide d'indicateurs de qualité comme « excellent » ou « insuffisant ». La définition de la performance idéale peut être tirée de normes comme les standards Sphère (Projet Sphère, 2011) ou de documents de projets ou encore des documents de politiques des organisations.

Les tableaux évaluatifs comportent deux éléments :

- Les descripteurs de performance – insuffisant, satisfaisant, bon, excellent, etc.
- Les critères qui définissent ce qui est considéré comme bon. Le modèle de Davidson présente cela sous la forme d'une liste de critères sous chacun des descripteurs de performance d'un sujet donné.

Performance	Excellent	Très bon	Bon	Satisfaisant	Insuffisant
<b>Descripteurs de performance</b>	Exemple évident de performance exemplaire ou de bonne pratique en matière de distribution alimentaire ; aucun point faible.	Très bonne ou excellente performance à presque tous les niveaux. Solide globalement mais pas exemplaire. Pas de point faible qui ait de réelle conséquence.	Performance générale assez bonne, peut présenter certains points faibles minimes mais rien de sérieux.	Performance correcte, quelques points faibles sérieux mais pas rédhibitoires.	Exemple évident d'un fonctionnement insatisfaisant ; des faiblesses sérieuses concernant systématiquement les aspects essentiels.
<b>Liste des caractéristiques de cette performance</b>					

Source : D'après Davidson (2004: 137)



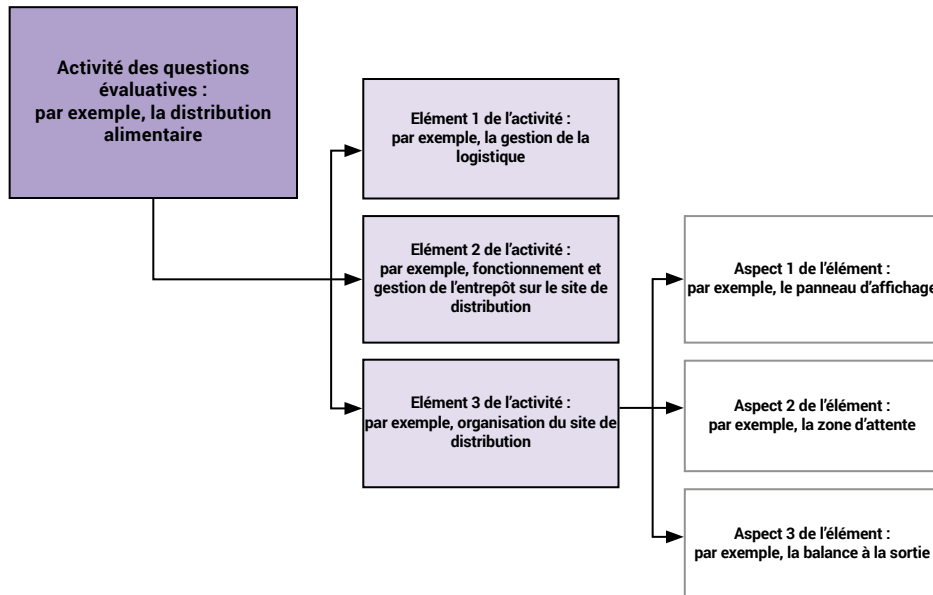
#### **Astuce :**

Concevez le tableau évaluatif avant la principale collecte de données. Définissez ce que vous pensez être les caractéristiques de différents niveaux de performance avant de commencer votre principale collecte de données. Ne réinventez pas la roue. Étudiez les modèles ou standards conceptuels du secteur pour concevoir les critères de votre tableau (Voir Section 5 : Elaborer votre évaluation).

Davidson utilise une liste de caractéristiques pour chaque descripteur de performance associé à chaque élément de l'activité, comme dans Oakden (2013: 6-7). Pour cet exemple, la liste de caractéristiques relatives aux excellentes performances de l'organisation pour un site de distribution alimentaire (un élément d'une activité de distribution alimentaire) peut inclure les aspects suivants :

- Panneau d'affichage illustré indiquant les échelles de rations pour chaque groupe dans les langues pertinentes
- Zone d'attente ombragée avec contrôle de l'accès et mise à disposition d'eau et de toilettes
- Balances d'assez grande capacité pour que les bénéficiaires puissent vérifier leurs rations – à l'aide de poids étalon
- Signalétiques claires dirigeant les personnes vers les bonnes files d'attente
- Files d'attente propres et sèches
- Cuillères en bon état et clairement marquées
- Dispositifs de contrôle des files d'attente pour éviter des bousculades

Globalement, cela revient à décomposer une activité plus complexe en éléments et en aspects pour permettre une évaluation plus transparente de la qualité de l'activité (déterminée par les critères exprimés dans [le tableau évaluatif](#)). Par exemple :



Le tableau évaluatif qui suit indique quelques aspects de l'organisation du site de distribution alimentaire :

<b>Activité</b>	Distribution alimentaire				
<b>Élément</b>	Organisation du site de distribution				
<b>Aspect</b>	<b>Excellent</b>	<b>Très bon</b>	<b>Bon</b>	<b>Satisfaisant</b>	<b>Insuffisant</b>
<b>Panneau d'affichage</b>	Panneau d'affichage illustré indiquant les échelles de rations pour chaque groupe dans les langues pertinentes	Panneau d'affichage sans illustrations indiquant les échelles de rations pour chaque groupe dans les langues pertinentes	Panneau d'affichage indiquant les échelles de rations pour chaque groupe mais certaines langues clés sont oubliées	Panneau d'affichage indiquant les échelles globales des rations mais certains groupes et certaines langues sont oubliés	Pas de panneau d'affichage
<b>Zone d'attente</b>	Zone d'attente ombragée avec contrôle de l'accès et mise à disposition d'eau et de toilettes	Zone d'attente ombragée avec contrôle de l'accès et mise à disposition de toilettes	Zone d'attente ombragée avec contrôle de l'accès mais pas de toilettes	Zone d'attente avec contrôle de l'accès	Pas de zone d'attente formelle
<b>Balances de pesée à la sortie</b>	Balances d'assez grande capacité pour que les bénéficiaires puissent vérifier leurs rations – à l'aide de poids étalon	Balances d'assez grande capacité pour que les bénéficiaires puissent vérifier leurs rations – mais pas de poids étalon	Balances de grande capacité pour que les bénéficiaires puissent vérifier leurs rations – mais files d'attente	Balances de pesée à la sortie mais pas assez grandes pour que les bénéficiaires puissent vérifier leurs rations complètes	Pas de balances de pesée à la sortie

Critères établis d'après le HCR (1997) ; Jaspers et Young (1995) ; Projet Sphère (2011).

Les tableaux évaluatifs peuvent utiliser des critères de performance quantitatifs autant que qualitatifs, et dans l'exemple suivant qui indique la notation vis-à-vis de la conformité à l'affirmation « les distributeurs ont traité les bénéficiaires avec respect », les notes obtenues en réponse à l'enquête sont utilisées pour définir dans quelle mesure les bénéficiaires ont été traités de façon digne.

<b>Catégorie</b>	<b>Excellent</b>	<b>Très bon</b>	<b>Bon</b>	<b>Satisfaisant</b>	<b>Insuffisant</b>
<b>Notes des aspects traités</b>	Plus de 90 % des personnes interrogées sont d'accord dans une large ou très large mesure	80-90 % sont d'accord dans une large ou très large mesure	60-80 % sont d'accord dans une large ou très large mesure et pas plus de 15 % sont d'accord de manière limitée ou très limitée	40-60% sont d'accord dans une large ou très large mesure et pas plus de 15 % sont d'accord de manière limitée ou très limitée	Moins de 40 % sont d'accord dans une large ou très large mesure

Source : D'après Oakden (2013: 10)

Les notes des différents aspects observés peuvent être combinées dans une note globale pour l'élément, comme indiqué ci-dessous.

Note globale de l'élément	Excellent	Très bon	Bon	Satisfaisant	Insuffisant
Notes des aspects traités	Tous les aspects ont une note excellente	Presque tous les aspects ont une note très bonne ou excellente ; aucune note satisfaisante ou insuffisante	La plupart des aspects ont une note bonne ou mieux que bonne ; pas de note insuffisante, plus de notes très bonnes ou excellentes que de notes satisfaisantes	La plupart des aspects ont une note satisfaisante ou mieux que satisfaisante, moins de 20 % des notes sont insuffisantes.	Plus de 20 % des notes des aspects observés sont insuffisantes

La même approche peut être utilisée pour combiner les notes des éléments dans un seul jugement évaluatif global pour l'activité.



**Astuce :**

Concevez, ou encouragez votre équipe à concevoir, des tableaux évaluatifs lors d'un atelier avec des parties prenantes.

Le fait d'organiser un atelier avec des parties prenantes pour définir un tableau évaluatif de mesure du respect des délais réduit la probabilité de désaccords à ce sujet par rapport à la solution où les évaluateurs ou le responsable de l'évaluation cherche(nt) de façon arbitraire à concevoir ce tableau.

Davidson (2014) fournit une excellente présentation générale du raisonnement évaluatif. Aucun exemple humanitaire d'application des tableaux évaluatifs n'a été trouvé durant la rédaction de ce guide, mais Oakden (2013) fournit un exemple de son utilisation. Les tableaux évaluatifs sont un outil puissant pour contester le problème constant des jugements évaluatifs dans l'EAH qui consiste à les rejeter en les considérant comme purement subjectifs.



**Astuce :**

Utilisez les tableaux évaluatifs, ou encouragez votre équipe à les utiliser, lorsque vous pensez que certaines parties prenantes peuvent contester les réponses aux questions évaluatives en les jugeant subjectives ou fondées sur des anecdotes ou des biais.

## 16.4 Analyser les éléments de preuve pour des questions causales

Les questions causales (voir Section 6 : Choisir les questions d'évaluation sur les types de questions) interrogent sur les liens de causalité entre interventions et résultats. La causalité est un problème dans l'EAH parce que la plupart des interventions comprennent de nombreux acteurs et que tout changement est influencé par de multiples facteurs, plutôt que par un seul. Ainsi, l'EAH renvoie généralement à la contribution plutôt qu'à l'attribution.

L'inférence causale est la pierre angulaire de l'évaluation d'impact.



### Définition : Inférence causale

L'inférence causale est l'établissement d'une relation entre une cause et un effet.

Comme le Remarque Cook *et al.*, il faut comprendre que « l'inférence causale, même lors d'expérimentations, est fondamentalement qualitative » (2002 : 6). Décider de ce que signifie la relation entre variables est un jugement qualitatif, même si cette action se fonde sur des données quantitatives. Par exemple, décider quels facteurs doivent être vérifiés en raison d'un potentiel risque de confusion est un jugement qualitatif.

Davidson (2009) propose huit stratégies pour établir une inférence causale :

1. Interrogez ceux qui ont observé ou vécu la causalité étudiée de façon directe. Dans la discussion entre Cook et Scriven sur la causalité dans les évaluations (Cook *et al.*, 2010), Scriven indique clairement que certaines personnes peuvent observer la causalité en « première main ». Cook remarque que ces personnes doivent mettre de côté les explications alternatives. Dans l'EAH, si vous observez que des vendeurs sur les marchés réduisent les prix des aliments dès que la distribution alimentaire est annoncée, vous auriez de bonnes raisons de penser que l'annonce est la cause de la réduction une fois que vous aurez parlé aux vendeurs pour rejeter toute explication alternative.
2. Vérifiez si le contenu de l'intervention correspond à la nature du produit. Par exemple, si vous avez constaté une réduction de la malnutrition sévère suite à la distribution de matelas, il serait idiot d'en conclure que la distribution a mené à cette réduction.
3. Recherchez des types d'effet distinctifs (la méthode de *modus operandi*) tout en cherchant et en éliminant des explications alternatives.
4. Vérifiez si le *timing* des résultats fait sens. Si B se produit systématiquement après A, alors il est probable que A puisse mener à B.
5. Observez la relation entre la « dose » et la « réponse ». Par exemple, est-ce que les enfants qui ont eu la même exposition à un événement

- traumatique présentent moins de symptômes s'ils ont bénéficié d'un plus grand nombre de sessions thérapeutiques que ceux qui en ont eu moins durant le même intervalle de temps.
6. Utilisez une comparaison ou un système de contrôle (conception expérimentale ou quasi-expérimentale).
  7. Utilisez des modèles statistiques pour éliminer d'autres facteurs potentiels de causalité. Cette approche est utilisée dans les études quasi-expérimentales fondées sur des statistiques qui utilisent des techniques comme la stratification, les modèles multivariés, la régression logistique et linéaire, et l'analyse de covariance pour exclure l'influence de variables qui ne vous intéressent pas.  
Voir Pourhosingholi *et al.* (2012) pour une introduction à ces techniques.
  8. Identifiez et vérifiez les mécanismes de causalité. Vous pouvez suivre la théorie du changement (TdC) et vérifier les éléments de preuve de chaque composante de la chaîne causale. Voir Funnell et Rogers (2011) pour plus d'informations sur les théories du changement et leur utilisation.

Voir Alexander et Bonino (2015) pour une analyse de la causalité dans l'EAH. Rogers (2014) fournit également une bonne introduction à la causalité. Enfin, Befani (2012) propose une analyse universitaire minutieuse au sein de l'ouvrage de l'étude de Stern *et al.* (2012).

## 16.5 Analyser les éléments de preuve pour des questions descriptives

Les questions descriptives (voir Section 6 : Choisir les questions d'évaluation) requièrent parfois l'utilisation de méthodes statistiques (voir Section 16.6) pour décrire l'intervention en identifiant notamment des moyennes. L'analyse de données qualitatives codées peut également être utilisée pour identifier les thèmes et problématiques globaux de l'intervention.

### L'analyse des données qualitatives relevant des données primaires

L'analyse des données qualitatives doit être un processus rigoureux et logique qui convertit les données de l'évaluation en résultats. L'analyse de données qualitatives formelles correspond généralement au fait de rattacher à des catégories des éléments de preuve précis et ensuite de réunir ces éléments. Le fait d'attribuer ces catégories est appelé « codage ».

**Définition : Codage**

Le codage attribue des catégories à des éléments de preuve précis.

Les catégories sont décrites par des codes, qui peuvent être simplement des mots-clés décrivant la catégorie ou, dans des types de codage plus complexes, une série de caractères alphanumériques. Dans l'EAH, le codage se limite presque toujours à l'utilisation de mots-clés qui restent le codeur de la catégorie.

**Le codage dans l'EAH**

Sous sa forme la plus simple, le codage peut consister à lire les notes de terrain et attribuer des étiquettes à différentes catégories d'éléments de preuve, ou simplement à enregistrer les notes de terrain selon les catégories définies au départ.

Voici quelques exemples d'approches simples :

- Parcourir les notes et attribuer des étiquettes (codes) aux éléments de preuve par rapport à la catégorie associée.
- Couper et coller les notes dans une nouvelle structuration organisée par thèmes.
- Utiliser des marqueurs ou des post-its de couleurs sur des documents physiques.
- Ajouter des étiquettes de texte, des commentaires ou des surlignements aux documents électroniques.

**Pour approfondir : Le codage dans l'EAH**

Les évaluations plus complexes requièrent des approches plus rigoureuses, comme le codage des notes de terrain. Dans l'EAH, les codes prennent généralement la forme de mots-clés et les codes complexes sont rarement utilisés. Un élément de preuve peut être une simple observation, une phrase ou une expression tirée d'un rapport ou d'un entretien, une réponse à une enquête, un résumé d'enquête ou tout autre exposé de faits ou opinion. En général, un entretien semi-structuré de 45 minutes peut produire de 0 à 10 ou 20 éléments de preuve. Saldaña (2012: 1-41) constitue une excellente introduction à ce sujet.

Les contraintes de budget et de temps signifient que le codage qui suit le travail de terrain est très inhabituel dans l'EAH. On peut citer comme exception l'évaluation CARE du Cambodia Disaster Preparedness and Mitigation où l'ensemble de l'équipe était engagée dans une « traduction et un codage des données qualitatives et dans l'écriture du rapport provisoire » (Ramage *et al.*, 2006). Dans l'EAH, le modèle habituel est le suivant : les membres de l'équipe d'évaluation écrivent des sections du rapport en s'appuyant sur leurs notes de terrain, ou alors le chef de l'équipe d'évaluation s'en charge à partir de notes écrites des membres de l'équipe et de ses propres notes.







Une bonne pratique actuelle de l'EAH consiste à coder les données qualitatives au fur et à mesure qu'elles sont collectées. Les codes initiaux sont déterminés par les questions d'évaluation et par les sujets qui émergent durant la phase de démarrage. Une évaluation humanitaire type peut avoir de 20 à 30 codes. Des codes supplémentaires peuvent être tirés (d'après Miles *et al.*, 2013) :

- Du cadre conceptuel (voir Section 5 : *Elaborer votre évaluation*). Par exemple, si l'approche des moyens de subsistance durables est utilisée comme cadre conceptuel, des codes seraient attribués pour la vulnérabilité, les différents actifs, les structures, les processus, les stratégies et résultats des moyens de subsistance.
- Des sujets à problèmes. Les sujets à problèmes courants dans l'action humanitaire sont le ciblage, le genre et la coordination.
- De l'hypothèse. Si les évaluateurs examinent l'hypothèse selon laquelle l'assistance en espèces a mené à des comportements antisociaux, alors ils devraient définir des codes pour différents types de comportements antisociaux. Des éléments de preuve liés à l'absence d'un tel comportement seraient également inclus.
- Des sujets clés. Dans le cadre d'un projet Abri, les sujets clés peuvent inclure les niveaux des loyers, si les abris sont habités par plusieurs familles, etc.

Le tableau de codes qui suit montre les codes utilisés par l'ETR inter-organisations de la crise dans la vallée de Swat (Cosgrave, 2009). Les codes découlaient d'une question d'évaluation, de sujets à problèmes, de questions clés, de points soulevés dans les TdR ou de thèmes ayant émergé durant le travail de terrain (comme la neutralité, l'accès, la bureaucratie, les retards et les facteurs contextuels). Plusieurs codes provenaient de sources multiples, mais seule la source principale est citée.

Code	Source du code	N° des éléments de preuve	N° des différentes sources
<b>Evolution entre 2009 et 2010</b>	Point des TdR	57	23
<b>Interdépendance</b>	Question d'évaluation	64	37
<b>Relèvement précoce</b>	Point des TdR	41	26
<b>Genre</b>	Sujet à problèmes	26	16



Code	Source du code	N° des éléments de preuve	N° des différentes sources
<b>Gouvernement</b>	Point des TdR	53	24
<b>Apprentissage pour l'avenir</b>	Point des TdR	61	24
<b>Militaires</b>	Question d'évaluation	35	16
<b>Suivi</b>	Point des TdR	23	16
<b>Neutralité</b>	Travail de terrain	92	36
<b>Etudes</b>	Question d'évaluation	7	5
<b>Réalisations</b>	Question d'évaluation	81	39
<b>Financement</b>	Questions clés	91	36
<b>Coordination</b>	Questions clés	144	39
<b>Accès</b>	Travail de terrain	32	21
<b>Changements en matière d'apprentissage</b>	Point des TdR	22	12
<b>Sécurité</b>	Sujet à problèmes	56	30
<b>Diagnostics</b>	Question d'évaluation	43	32
<b>Camps-Hôtes-Autre</b>	Point des TdR	58	34
<b>Consultation</b>	Question d'évaluation	12	11
<b>Lacunes</b>	Sujet à problèmes	109	47
<b>Ciblage</b>	Sujet à problèmes	82	45
<b>Bureaucratie</b>	Travail de terrain	38	21
<b>Retards</b>	Travail de terrain	18	14
<b>Problèmes liés aux contextes</b>	Travail de terrain	57	29
<b>Total</b>		<b>1 306</b>	<b>160 au total</b>





Le codage est effectué en examinant les données collectées et en extrayant des éléments de preuve :

1. Les notes des entretiens et autres : l'équipe d'évaluation saisit les notes des entretiens dans le tableau des observations et les code en même temps.
2. Les documents papier : les passages clés sont retranscrits dans le tableau des observations.
3. Les documents électroniques : les passages clés sont copiés, nettoyés et collés dans le tableau des observations. (Un nettoyage est nécessaire pour enlever les retours à la ligne et les codes de formatage du texte.)
4. Les rapports d'enquête : pour les enquêtes en ligne, le processus peut être automatisé si les questions de l'enquête sont alignées sur les codes. Le résultat est vérifié manuellement et des codes supplémentaires sont ajoutés selon les besoins.

Le codage est bien plus aisé si tous les outils de collecte des données, les guides d'entretien, les formulaires d'enquête, etc., sont alignés sur les codes de départ.



**Astuce :**

Utilisez les catégories de départ pour organiser toute la collecte des données, parce que si les catégories changent plus tard, il sera plus facile de les modifier à partir de données organisées que de données désorganisées.

Utilisez un tableau des observations simples avec feuilles de calcul pour faciliter le processus de codage : il pourra être consolidé automatiquement de façon à présenter un seul résumé des éléments de preuve.

Code	Éléments de preuve	Source	Initiales	2° Code	3° Code
<b>Le code du sujet lié à cet élément de preuve</b>	Détails concernant l'élément de preuve – en général, un extrait de texte – d'une phrase à un paragraphe. Le nombre de mots moyen d'un élément de preuve dans l'exemple donné était de 22.	Un code renvoyant à la source précise, que ce soit un entretien, une réunion, un document ou une observation	Qui l'a rédigé	Un deuxième code pour le même texte	Un troisième code pour le même texte



Même si ce tableau peut être utilisé avec un seul code pour chaque élément de preuve, des codes supplémentaires peuvent être ajoutés si nécessaire pour le même élément de preuve. Si plus de trois codes s'appliquent au même élément de preuve, le texte peut être ressaisi avec les codes supplémentaires.

Les sources des éléments de preuve peuvent être variées. Dans le cas d'une ETR aux Philippines, plus de 40 % des éléments de preuve venaient de commentaires précis formulés en réponse à des questions ouvertes d'une enquête en ligne (Cosgrave et Ticao, 2014). Cela, combiné avec des entretiens téléphoniques, a joué un rôle important dans l'évaluation car de nombreux informateurs clés primordiaux avaient déjà quitté les Philippines.

Méthode de collecte des données	Éléments de preuve	%
Enquête en ligne	944	42,1%
Entretien téléphonique	523	23,4%
Entretien semi-structuré (deux personnes interrogées ou plus)	390	17,4%
Entretien semi-structuré (une seule personne interrogée)	230	10,3%
Réunion collective	117	5,2%
Commentaires par e-mail	21	0,9%
Discussion approfondie (> 10 minutes sur un ou plusieurs sujets)	11	0,5%
Observation	2	0,1%
<b>Nombre total d'éléments de preuve</b>	<b>2 238</b>	<b>100%</b>

### Tirer des conclusions sans utiliser le logiciel CAQDAS

L'équipe d'évaluation utilise les codes pour trier toutes les données et les rassembler par code. Le chef d'équipe ou un membre de l'équipe affecté à certains codes particuliers passe ensuite en revue toutes les données collectées pour un sujet spécifique avant d'en tirer des conclusions. Au moment de traiter les données codées, certains conflits apparaissent de façon inévitable.

Les conflits entre éléments de preuve peuvent se produire avec des données quantitatives tout autant qu'avec des données qualitatives. Un exemple de conflit entre éléments de preuve quantitatifs est celui où les quantités distribuées par mois

ne correspondent pas aux quantités mensuelles sorties des entrepôts – des raisons complexes peuvent expliquer cela, notamment des sorties et des distributions à des mois différents, des définitions différentes, du stockage local aux sites de distribution, la manière de gérer les anomalies aux sites de distribution, les retours ou les pertes, etc.

Lorsque vous prenez en compte les conflits entre éléments de preuve :

- Donnez plus de poids au point de vue des personnes les plus affectées. Par exemple, le point de vue des bénéficiaires sur le fait qu'ils aient ou non été consultés pèse plus que celui du personnel de l'organisation sur le même sujet. Le point de vue des bénéficiaires qui n'ont pas reçu toute l'aide en raison de lacunes de coordination doit peser plus lourd que celui des personnes qui ont suivi les réunions de coordination.
- Le point de vue d'une personne ayant une expérience importante d'un secteur peut avoir plus de poids que celui d'une personne peu expérimentée. Ce n'est pas toujours le cas toutefois : une personne dotée d'une solide expérience peut être enfermée dans une approche ou un état d'esprit particulier.
- Envisagez de donner plus de poids aux opinions qui semblent aller à l'encontre de l'intérêt de la personne interrogée. Par exemple, soyez très attentif aux bénéficiaires qui disent ne pas vouloir davantage d'un certain type d'assistance. De même, soyez très attentif lorsque le personnel d'une organisation critique son propre programme, en particulier s'il ne fournit pas d'autres signes de désaffection.
- Donnez plus de poids aux points de vue qui se triangulent avec d'autres éléments de preuve.

Cette approche garantit que les conclusions du rapport d'évaluation sont fortement enracinées dans les éléments de preuve rassemblés. Là où des éléments de preuve font l'objet de conflits importants, ceux-ci doivent être signalés de même que le raisonnement qui a été utilisé pour les estimer.

### Le logiciel CAQDAS pour le codage post-terrain

Le codage post-terrain peut être un processus lent et il a rarement lieu dans l'EAH. Il existe toutefois des logiciels puissants qui peuvent aider à cela, comme le Computer Assisted Qualitative Data Analysis (CAQDAS). Néanmoins :

- Ils coûtent chers
- Le codage est très chronophage
- La prise en main des logiciels n'est pas rapide

A cause de cela, les logiciels CAQDAS sont rarement utilisés dans l'EAH. LaPelle (2004) montre qu'un logiciel de traitement de texte d'utilisation générale peut être utilisé pour des tâches CAQDAS de base. Koenig (2004) constitue une très bonne introduction au logiciel CAQDAS, et notamment aux potentiels écueils d'utilisation.

Miles *et al.* (2013) note que « Certains programmes CAQDAS comprennent des fonctionnalités très utiles pour soutenir les analyses, mais la plupart des programmes ont une courbe d'apprentissage abrupte. » Silver et Lewins (2014) proposent un guide par étape pour certains des logiciels les plus utilisés. Néanmoins, Baugh *et al.* (2010) décrivent le temps nécessaire à l'apprentissage d'un logiciel comme l'une des plus grandes limites à l'utilisation du CAQDAS.

## L'analyse des données qualitatives issues de données secondaires comme les rapports d'étapes

Ce sujet a déjà été traité en partie dans la Section portant sur les [Section 10 : Les méthodes de recherche documentaire](#). Ici, l'objectif est de souligner que les données secondaires peuvent également être intégrées dans le tableau des observations. Le plus important travail de terrain de l'évaluation peut permettre d'accéder à des données secondaires complémentaires qui devront être analysées en utilisant les méthodes décrites pour les revues documentaires.

### L'analyse de portefeuille

L'analyse de portefeuille est une technique puissante pour examiner la stratégie en observant les données d'un projet. Sous sa forme la plus simple, l'analyse de portefeuille peut simplement consister à analyser de façon descriptive des sujets comme le pourcentage de l'aide qui est allé à quels partenaires ou secteurs, etc. Elle peut également examiner le montant moyen des différents types de subvention. C'est le type d'analyse de portefeuille réalisée par l'évaluation de la stratégie humanitaire danoise (Mowjee *et al.*, 2015) et par l'évaluation conjointe du soutien au maintien de la paix au Soudan du Sud (Bennett *et al.*, 2010).

A un niveau légèrement plus détaillé, elle peut observer quels types de projet, secteurs ou partenaires étaient les plus à même de ne pas demander une extension budgétaire ou de ne pas rencontrer de problèmes.

L'analyse de portefeuille peut également impliquer d'utiliser un outil de notation ou un tableau de synthèse, comme dans le cas de l'évaluation australienne de l'aide fournie par les ONG suite au séisme au Pakistan d'octobre 2005 (Crawford *et al.*, 2006). Elle peut également impliquer la notation de projets particuliers.

Une analyse de portefeuille plus approfondie peut consister à mettre en place un certain nombre d'outils de notation et ensuite à noter chaque projet du portefeuille en utilisant ces derniers. C'est l'approche qui a été retenue par l'évaluation des cinq années du CERF (Channel Research, 2011) en ce qui concerne les domaines du genre, de la vulnérabilité et les autres sujets transversaux.

## 16.6 L'analyse statistique des données primaires

La présentation qui suit de l'analyse statistique n'est pas exhaustive. Ce sujet est très technique et de nombreux documents le traitent, tous sont beaucoup plus longs que ce *Guide*.

Il est à noter que les données collectées dans les enquêtes ne sont pas seulement numériques et qu'elles peuvent également comprendre des données catégorielles. Les quatre catégories de données sont :

- Catégorielles
  - Nominales – comme le genre (femmes/hommes)
  - Ordinales – où les catégories sont classées (classe sociale, par exemple, ou « très insatisfait, insatisfait, satisfait et très satisfait » en réponse à une question d'enquête)
- Numériques
  - Données d'intervalle – données numériques pour lesquelles il n'y a pas de zéro absolu (comme pour une note de sécurité alimentaire)
  - Données de ratio – données numériques où un score de zéro a du sens et où les nombres peuvent être multipliés en gardant du sens (comme le gain de poids dans un programme de nutrition)

Les procédures statistiques qui peuvent être utilisées varient avec le type de données.

### Nettoyage et saisie des données

L'utilisation d'une collecte des données électronique simplifie grandement le nettoyage et la saisie des données. Celles-ci ont encore besoin d'être vérifiées car le logiciel ne corrigera pas toutes les erreurs.

La collecte des données électronique a fortement remplacé ce qu'était la tâche onéreuse de saisie des données.

### Données manquantes

L'absence de données (provenant de non-réponses ou de réponses partielles) pose le problème suivant : on craint toujours que ceux qui n'ont pas répondu soient d'une manière ou d'une autre différents de ceux qui ont répondu, et que cela puisse biaiser les résultats. C'est une préoccupation réelle, en particulier avec les enquêtes en ligne, même si la non-réponse est également un problème dans le cadre des enquêtes en face-à-face. Les raisons qui expliquent l'absence de réponse sont notamment :

- La personne à interroger n'était pas disponible (l'enquête peut inclure un protocole pour renouveler ces tentatives d'entretien ou pour prévoir des informateurs de remplacement)
- Le refus de répondre à des questions précises
- Incapacité à répondre à des questions précises (ne sait pas)
- Oubli de l'enquêteur

Une bonne conception et un bon test du questionnaire peuvent aider à réduire les trois dernières catégories. La première demeure un sujet de préoccupation particulier parce qu'une personne potentiellement à interroger peut être absente pour des raisons qui biaisent les résultats de l'enquête. Par exemple, si une enquête a constaté que 10% de l'échantillon sélectionné n'était pas disponible lorsque les enquêteurs ont appelé, la raison peut être que ces personnes étaient engagées dans des activités importantes en dehors du village. Les exclure pourrait biaiser l'enquête car cela pourrait exclure un groupe présentant un moyen de subsistance spécifique. Encore une fois, les tests en amont peuvent identifier des problèmes de ce genre.

## Les statistiques

L'utilisation de statistiques dans l'EHA est de deux types : descriptifs et inférentiels.



### Définition : Statistiques descriptives

Les statistiques descriptives sont utilisées pour résumer des aspects essentiels d'une population.

Les statistiques descriptives les plus courantes sont des mesures de tendances centrales (communément appelées moyennes) et des mesures d'écart ou de répartition (voir p. 313). De telles statistiques sont souvent utilisées pour donner une réponse partielle aux questions d'évaluation descriptives. Elles sont également utiles pour répondre à des questions normatives, en particulier lorsque la norme ou le standard en question possède un indicateur chiffré. Par exemple, la réponse à une question comme « Dans quelle mesure avons-nous respecté les standards Sphère pour la quantité d'eau fournie ? » devrait probablement comprendre des données sur la quantité moyenne fournie par personne et par jour, ainsi que la fourchette d'estimation avant une discussion sur le fait de savoir si ce niveau d'assistance a respecté le standard correspondant à une « quantité suffisante d'eau de boisson, de cuisson, d'hygiène domestique et personnelle ».



### Définition : Statistiques inférentielles

Les statistiques inférentielles sont utilisées soit pour tirer des déductions sur une population à partir d'un échantillon, soit pour tirer des déductions sur des hypothèses.

Les statistiques inférentielles sont utilisées à la fois pour répondre à des questions descriptives (comme : quel pourcentage de la population affectée a utilisé la plus grande part de son aide en espèces pour des frais d'abri ?) et des questions causales (comme : dans quelle mesure le programme communautaire de malnutrition aiguë a-t-il réduit les niveaux de malnutrition sévère et aiguë ?).





### Astuce :

Indiquez toujours le niveau et l'intervalle de confiance des statistiques inférentielles.

Par exemple, si 50 % de l'échantillon a utilisé la plus grande part de son aide en espèces pour un abri, mais qu'en raison de la petite taille de l'échantillon l'intervalle pour le niveau de confiance de 95 % était de 20 % à 80 %, c'est un fait très pertinent puisqu'on est à 95 % certain qu'entre 20 % et 80 % de la population a utilisé la plus grande part de son aide en espèces pour un abri.

### La tendance centrale

La mesure la plus courante de tendance centrale pour des données numériques est probablement la moyenne comme, par exemple, la taille de la famille moyenne ou le temps moyen passé dans un camp de transit. La tendance centrale est souvent utilisée pour répondre à des questions descriptives, et elle peut décrire notamment le nombre moyen ou le volume moyen d'aide en espèces distribuées. Utiliser les moyennes de cette manière peut rapidement donner une idée de l'échelle des interventions.

La moyenne est une mesure unique de la tendance centrale. Les deux autres mesures possibles de la tendance centrale sont la médiane, qui est la valeur pour laquelle la moitié des valeurs sont élevées et la moitié sont plus petites, et le mode, qui est la valeur la plus courante. La valeur médiane peut souvent être une mesure plus utile de la tendance centrale.

### De nouveaux réfugiés dans un camp de réfugiés déjà en place

Prenez un camp de 2.000 réfugiés qui sont là depuis 20 ans. Une nouvelle arrivée de 8.000 réfugiés a cherché refuge dans le camp il y a un mois. La durée moyenne du séjour des réfugiés dans le camp est de  $(2.000 \times 20 + 8.000) / 10.000 = 48,8$  mois, soit plus de quatre ans. Cependant, la médiane est de seulement un mois (si vous alignez tous les réfugiés par date d'arrivée, les numéros 4.999 et 5.000 auraient été là depuis un mois). Le mode (la durée de séjour la plus courante dans le camp) est également d'un mois.

### La dispersion

Les mesures de dispersion (ou de répartition) montrent comment les valeurs sont réparties autour de la moyenne ou de la médiane et peuvent être très importantes pour répondre à des questions descriptives. Par exemple, il ne serait pas très pertinent d'indiquer que l'aide en espèces moyenne pour les activités de moyens de subsistance était de 300 dollars sans mentionner également qu'elles allaient de 30 \$ à 600 \$, mais

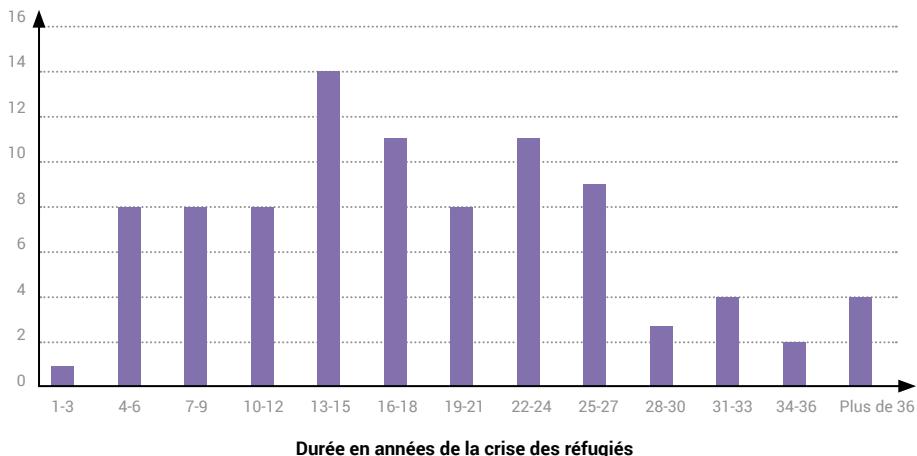
que 90 % des bénéficiaires recevaient entre 200 \$ et 400 \$. Voici quelques types de mesures de dispersion :

- La fourchette d'estimation, c'est-à-dire la différence entre les valeurs les plus hautes et les plus basses (par exemple, la taille des familles s'échelonnant de 1 membre à 19).
- Les sous-gammes, la différence entre les valeurs les plus hautes et les plus basses qui s'appliquent à un groupe spécifique (les ménages dirigés par une seule personne ont reçu une aide comprise entre 150 \$ et 450 \$).
- L'écart interquartile, c'est-à-dire la différence entre les valeurs qui sont de 25 % et de 75 % des valeurs notées.
- La variance – la somme des carrés de la différence entre chaque valeur et la moyenne.
- L'écart-type – la racine carrée de la variance. Cette mesure a l'avantage d'être dans la même unité que la moyenne : par conséquent, si la mesure se fait en années, l'écart-type devra également être en années.

Cependant, la variance et l'écart-type sont des mesures assez techniques. Les tableaux de répartition sont souvent plus informatifs pour les principaux utilisateurs visés. Le tableau qui suit présente les informations de façon plus simple que ne l'aurait fait un constat de la durée moyenne des déplacements et de l'écart-type.

**Schéma 16.2** : Distribution statistique des crises de réfugiés dans 91 pays d'origine entre 1978 et 2014 comprenant plus de 5.000 réfugiés à leur niveau maximal (hormis les récentes crises en Syrie et au Soudan du Sud)

#### Nombre de pays



Source : Elaboré par l'auteur

## Les tests d'hypothèses statistiques

Le principal avantage de la collecte de données quantitatives est de permettre de tester des hypothèses en utilisant des statistiques. Voici des types d'hypothèses :

- Les hypothèses portant sur les caractéristiques d'un seul groupe à partir des caractéristiques d'un échantillon raisonné (par exemple, estimer la population moyenne).
- Les hypothèses portant sur la différence ou la relation entre deux groupes (par exemple, entre le groupe assisté et un groupe de comparaison qui n'a pas reçu l'assistance, ou entre le groupe assisté avant et après l'intervention).

L'approche habituelle consiste à tester la différence entre l'hypothèse et l'hypothèse nulle. Si, par exemple, cette hypothèse est que ceux qui ont reçu un kit abri il y a un an avaient un meilleur résultat de sécurité alimentaire au niveau du ménage qu'un groupe équivalent qui ne l'a pas reçu, alors l'hypothèse nulle serait qu'ils n'ont pas un meilleur résultat de sécurité alimentaire au niveau du ménage.

L'hypothèse nulle est semblable au scénario contrefactuel, qui est la condition qui aurait prévalu s'il n'y avait pas eu d'assistance. Dans cet exemple, le groupe équivalent qui n'a pas reçu le kit abri représente le scénario contrefactuel. Il n'est jamais possible d'observer réellement le scénario contrefactuel (parce qu'il est impossible de tester simultanément les résultats de l'assistance et de la non-assistance, c'est-à-dire au même moment pour la même personne, si bien que l'hypothèse nulle est une approximation).

En testant une hypothèse, il est possible de conclure soit que l'hypothèse est vraie, soit qu'elle est fausse. Si nous concluons qu'elle est vraie, on peut avoir raison ou tort. Si nous avons tort en concluant que l'hypothèse est fausse, c'est une fausse erreur positive et nous utilisons dans ce cas une valeur de 5 % par convention : en d'autres termes, notre conclusion que l'hypothèse est vraie a une probabilité de 5 % d'être fausse.

Si nous concluons à tort que l'hypothèse est fausse, il s'agit d'une erreur négative fausse. Par convention, on lui attribue une valeur de 20 %, mais nous renvoyons à la [Section 12 : L'échantillonnage](#) où l'on laisse entendre qu'une probabilité de 10 % d'erreur négative fausse est plus appropriée.

## Les tests statistiques

Les tests utilisés dépendent du type de données (catégorielles ou numériques, comme indiqué ci-dessus) et du nombre de variables dépendantes et indépendantes.

- Variable dépendante : le résultat évalué – dans cet exemple, les scores de la sécurité alimentaire des ménages. C'est une variable numérique.
- Variable indépendante : les ressources qui ont un intérêt pour l'évaluation – dans ce cas, la fourniture de kits abri. C'est une variable nominale (ont reçu le kit/ n'ont pas reçu le kit).

Le choix d'un test approprié peut être un problème technique complexe.

Les tests statistiques précis dépendront également de la conception de l'évaluation. Différentes analyses sont appropriées pour des régressions multiples, pour des séries chronologiques interrompues, pour une régression discontinue, etc. Les équipes d'évaluation peuvent réaliser de telles conceptions seulement si elles ont les compétences techniques pour les analyser.

En règle générale, il faut tester la signification statistique de tous les résultats apparents. Dans ce cas, par exemple, si nous avons constaté que le groupe assisté a eu un résultat de sécurité alimentaire des ménages était 5 % plus élevé que celui du groupe non-assisté, il faut tester la probabilité que ce résultat soit dû au hasard.

Pour plus de détails sur la réalisation de tests statistiques, veuillez consulter des documents fiables. Actuellement, de nombreux textes universitaires portant sur les statistiques semblent s'appuyer sur l'utilisation d'un logiciel comme SPSS.<sup>17</sup> De telles solutions informatiques sont assez onéreuses. Microsoft Excel offre des possibilités statistiques de base qui peuvent être améliorées avec une extension gratuite comme Real Statistics (Zaointz, 2015) qui offre toute une gamme de fonctions, notamment certaines utiles pour traiter des données quantitatives.

## 16.7 L'analyse numérique des données secondaires, comme celles issues des registres de données de distributions

L'analyse numérique des données secondaires est semblable à l'analyse des données primaires, mais elle se concentre davantage sur les statistiques descriptives, les tendances et les corrélations. L'étude des déplacements de longue durée (Crawford *et al.*, 2015) a été partiellement élaborée autour de l'analyse des données secondaires collectées par le HCR sur les réfugiés, mais elle a également utilisé des données secondaires sur les PDI provenant du Internal Displacement Monitoring Centre, des données du développement provenant du PNUD, des données économiques de la Banque mondiale, des données en matière de stabilité provenant du Fragile States Index, ainsi que des données financières du HCR, du PAM et de l'UNICEF.

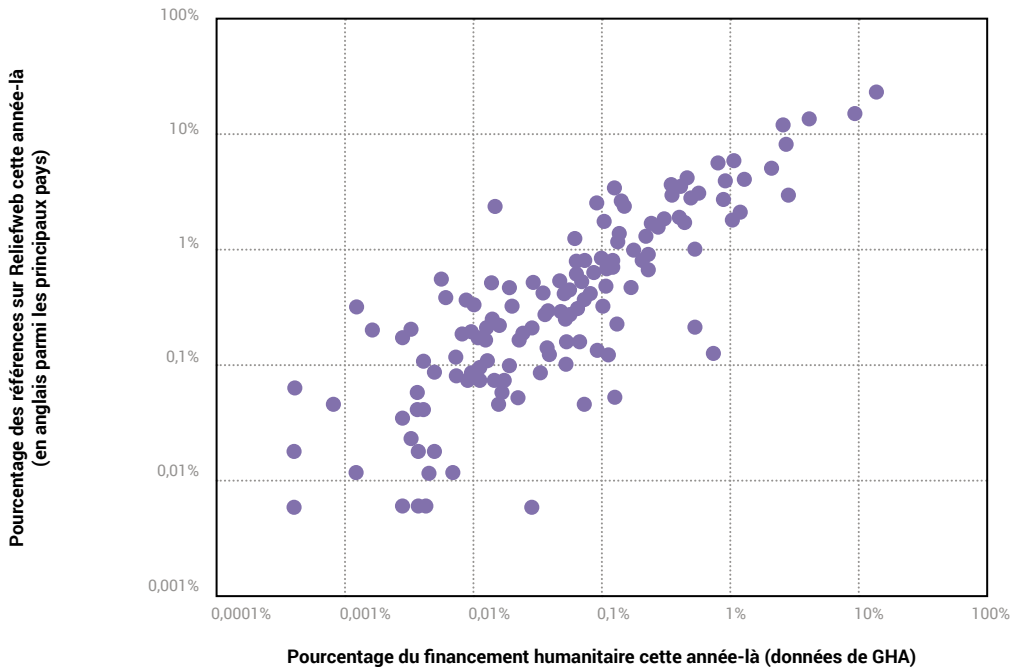


**Astuce :**

Créez toujours un diagramme de dispersion pour toutes les données numériques que vous étudiez.

Les diagrammes de dispersion font ressortir toutes les anomalies, montrent dans quelle mesure les données sont corrélées et peuvent déboucher sur des secteurs d'enquête utiles. Tout tableau peut produire un diagramme de dispersion. Il est important de se rappeler que la corrélation n'implique pas la causalité. Il est possible que la corrélation soit entièrement due au hasard ou que les variables apparemment dépendantes et indépendantes soient en fait influencées par une troisième variable ou plus.

**Schéma 16.3 :** Corrélation entre le financement humanitaire global et les références sur ReliefWeb en 2010 (chaque point représente un pays)



Source : Elaboré par l'auteur

L'étude des déplacements de longue durée (Crawford *et al.*, 2015) a utilisé des diagrammes de dispersion pour tester les relations entre données avant de réaliser un test statistique. Dans ce cas, le diagramme de dispersion testait la relation entre le financement humanitaire et les références à la crise sur ReliefWeb par pays donné. Le premier objet d'étude cherchait à savoir si la répartition des données ReliefWeb pouvait être utilisée comme indicateur indirect pour la répartition des financements humanitaires en ce qui concerne les dernières années où les données financières ne sont pas encore disponibles.

## Les tests statistiques couramment utilisés dans l'analyse numérique

Toute équipe d'évaluation doit être capable de réaliser des tests statistiques pour :

- Vérifier la corrélation. Quelle est la probabilité qu'une corrélation se soit produite par hasard ?
- Des tests d'indépendance, comme les tests du chi carrés. Deux facteurs sont-ils liés de quelque manière que ce soit ou indépendants ?

Un tableur courant comprend les fonctions nécessaires pour réaliser de tels tests. De tels tests statistiques de base sont essentiels pour déterminer si une corrélation est susceptible de s'être produite par hasard. Les tests d'indépendance sont nécessaires pour déterminer si deux facteurs sont indépendants avant de perdre du temps à approfondir leur relation.

Par exemple, lors de l'évaluation du Consortium des organisations humanitaires britanniques (Cosgrave et Polastro, 2012), l'équipe a utilisé un test du chi carré pour vérifier la relation entre le fait d'être membre du comité des subventions et le pourcentage des demandes de subvention ayant obtenu un financement. L'impact du statut de membre du comité sur le succès de la demande ne s'est pas révélé significatif sur le plan statistique car il seulement atteint le niveau de 5 %.

## 16.8 Des éléments de preuve aux recommandations

Les rapports d'évaluation comprennent généralement des éléments de preuve, des résultats, des conclusions et des recommandations. Les éléments de preuve sont habituellement présentés pour illustrer comment l'équipe a obtenu ses conclusions à partir d'un résultat précis. La somme des éléments de preuve collectés utilisés dans le cadre d'une évaluation comprend les notes de l'équipe, toutes les données collectées et l'ensemble de références collectées. Seule une fraction de ces informations peut être présentée dans le rapport.

**Astuce :**

Donnez du poids aux éléments de preuve présentés dans le rapport. Insérez des citations de bénéficiaires et utilisez des photographies, des tableaux et des diagrammes pour illustrer et résumer des points clés, voir la [Section 14](#) pour en savoir plus sur ce sujet.

**Définition : Constatation**

Une constatation est une affirmation factuelle fondée sur un élément de preuve.

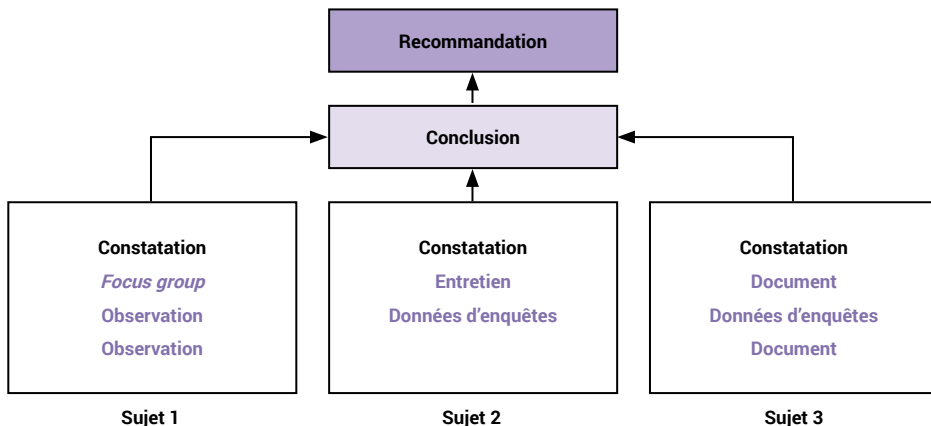
**Définition : Conclusion**

Une conclusion est une affirmation inductive fondée sur un ou plusieurs résultats.

**Définition : Recommandation**

Une recommandation est une façon de procéder que les évaluateurs suggèrent comme manière de traiter une ou plusieurs conclusions.

Il est nécessaire que le déroulement logique ou le processus d'analyse soit clair, depuis les constatations jusqu'aux recommandations en passant par les conclusions. Il doit être présenté de façon aussi transparente que possible dans les produits de l'évaluation. Par exemple, les constatations doivent clairement s'appuyer sur les éléments de preuve, les conclusions sur les constatations, et les recommandations sur les conclusions. Ce découpage respecte globalement la structure d'un rapport scientifique.



Jane Davidson (2007) plaide pour une autre physionomie des rapports d'évaluation. Elle suggère en effet que les réponses aux questions d'évaluation soient présentées en premier, suivies par les éléments de preuve et les conclusions. Elle plaide également en faveur d'une structuration du rapport autour des questions d'évaluation et de l'intégration de tout ce qui est lié à chaque question dans la section pertinente, notamment toutes les recommandations.

Les conclusions peuvent être présentées soit à la fin de chaque section, soit dans la section finale. Présenter les conclusions à la fin de chaque section les maintient plus proches des constatations associées et aide à rendre le déroulement logique plus transparent.

Il est également possible d'avoir une section finale qui résume les conclusions.

## Les recommandations

Les recommandations peuvent être :

- Omises du rapport. La logique qui se trouve derrière le fait de seulement tirer les conclusions consiste à penser que l'organisation commanditaire est la mieux placée pour transformer les conclusions en recommandations. Cette approche est rare dans l'EAH car très peu d'organisations disposent des structures nécessaires pour traduire les conclusions en recommandations. Le rapport final de l'ETR de la réponse de Plan International au cyclone tropical Haiyan aux Philippines ne comprenait pas de recommandations (Cosgrave et Ticao, 2014).
- Elaborées par l'équipe d'évaluation et présentées dans le rapport d'évaluation. C'est l'approche la plus courante dans l'EAH, et la plupart des rapports d'évaluation de l'EAH présentent des recommandations.
- Développées grâce à un processus consultatif impliquant les parties prenantes. Cette approche est tout aussi rare dans l'EAH. Les principes qui se trouvent derrière cette approche sont l'encouragement de l'appropriation des recommandations, et le fait que les parties prenantes sont les mieux placées pour suggérer des remèdes à des problèmes identifiés dans les conclusions. L'évaluation conjointe du séisme Yogyakarta (Wilson *et al.*, 2007) a utilisé cette approche (voir [exemple de bonnes pratiques à la page 346](#)).
- Affinées grâce à un processus consultatif impliquant les parties prenantes à partir des recommandations provisoires développées par l'équipe d'évaluation. Ce fut l'approche retenue par l'ETR des inondations au Pakistan en 2010 (Polastro *et al.*, 2011) (voir [exemple de bonnes pratiques à la page 346](#)). La politique d'ACF demande à ce que les recommandations soient générales pour que les parties prenantes puissent les développer (ACF, 2011).



## Les recommandations utiles

Les recommandations sont utiles seulement si elles sont :

- **Spécifiques** : ce qui est recommandé doit être extrêmement clair.
- **Liées à des actions vérifiables** : il doit être possible de dire si la recommandation a été mise en œuvre ou non.
- **Dirigées** : la personne ou l'entité responsable de la mise en œuvre de la recommandation doit être identifiée ; la responsabilité peut être davantage clarifiée dans la réponse de la direction au rapport.
- **Faisables** : les recommandations peuvent impliquer des manières de travailler innovantes ou inhabituelles mais elles doivent garder à l'esprit les ressources et autres contraintes.
- **Limitées dans le temps** : un calendrier de mise en œuvre des recommandations doit être donné quand cela est possible.
- **Cohérentes** : les recommandations ne doivent pas se contredire ou sembler se contredire.
- **Priorisées** : clarifier quelles sont les recommandations de premier et de second ordre.
- **Economiques** : les actions recommandées doivent clairement apporter des bénéfices proportionnels à leurs coûts.

Les recommandations doivent également être limitées en nombre. Moins un rapport contient de recommandations, plus l'organisation commanditaire peut l'utiliser facilement. Cependant, un évaluateur expérimenté peut remarquer des douzaines de problèmes de performance et a le devoir éthique de les signaler. Diverses stratégies peuvent aider à résoudre ce dilemme :

- Proposer une recommandation générale et indiquer ensuite en détail comment elle peut être mise en œuvre par différents acteurs.
- Emettre les recommandations mineures à l'oral et les placer dans l'annexe du rapport.
- Classer les recommandations par importance.
- Regrouper les recommandations par cible.

## Notes de fin de texte

### 11 / Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation

1. Le chapitre pertinent est disponible en ligne à l'adresse suivante : [us.sagepub.com/sites/default/files/upm-binaries/61527\\_Chapter\\_15.pdf](http://us.sagepub.com/sites/default/files/upm-binaries/61527_Chapter_15.pdf).
2. L'insu n'est pas toujours efficace, mais les études le testent rarement (Hrobjartsson *et al.*, 2007 ; Fergusson *et al.*, 2004). Sackett (2007) remarque que, dans trois études portant sur l'efficacité de l'insu, il n'est une réussite qu'une fois sur deux.

### 12 / L'échantillonnage

3. L'écart standard est la racine carrée de la somme des carrés des différences entre les valeurs individuelles d'une variable et la valeur moyenne de cette variable. L'écart standard est exprimé dans les mêmes unités que la variable.

### 13 / Les méthodes de terrain

4. Rédigé en partie d'après la page Internet de l'Université Michigan State consacrée aux questions d'approfondissement en entretien (Kennedy, 2006).
5. Grille comprenant le numéro de l'entretien sur un axe et le nombre d'adultes du ménage sur l'autre axe. L'intersection des deux donne un numéro aléatoire qui permet de sélectionner la personne interrogée (généralement en par ordre d'âge croissant).
6. L'étude NRC sur l'éducation en Palestine (Shah, 2014: 72) a utilisé un outil d'observation structuré pour observer les salles de classe. Cet outil comportait 52 entrées d'observation pour mesurer cinq concepts sous-jacents.
7. Voir [www.mande.co.uk/docs/hierarch.htm](http://www.mande.co.uk/docs/hierarch.htm).
8. Voir [betterevaluation.org/evaluation-options/richpictures](http://betterevaluation.org/evaluation-options/richpictures).



## Notes de fin de texte

### 14 / Impliquer la population affectée dans votre évaluation

9. Voir le réseau CDAC.
10. Près de 75 % des EAH évaluées par ALNAP entre 2001 et 2004 ont été jugées insatisfaisantes ou passables pour ce qui est de consulter les principales parties prenantes et d'encourager la participation, notamment les membres de la population affectée (Beck et Buchanan-Smith, 2008).
11. On peut trouver des conseils pour ce type d'évaluation participative dans le travail de Better Evaluation [betterevaluation.org/plan/approach/participatory\\_evaluation](http://betterevaluation.org/plan/approach/participatory_evaluation), ainsi que d'Alexander et Bonino (2014).
12. Voir Buchanan-Smith *et al.* (2015).
13. Cette sous-section s'inspire beaucoup du *Evaluation of Protection companion Guide* (Bonino, 2016). Le contenu a été adapté ici pour renvoyer davantage à l'EAH. Pour toute considération éthique spécifique en lien avec l'évaluation des programmes de protection, nous vous renvoyons à ce *Guide*.

### 15 / L'accès restreint

14. Pour une analyse détaillée de ce qui est souvent désigné comme « la réduction de l'espace humanitaire », voir Collinson et Elhawary (2012).
15. Voir également Bush et Colleen (2015).

### 16 / L'analyse

16. Le Projet Sphère remplacera les six standards courants par les neuf engagements et critères de qualité de la norme humanitaire fondamentale (CHS Alliance, 2015) dans les prochaines versions du Manuel Sphère (Projet Sphère, 2015).
17. Andy Field (2006) en est un exemple.

## Notes

11 / Les conceptions de l'évaluation permettant de répondre aux questions d'évaluation	192 - 213
12 / L'échantillonnage	214 - 230
13 / Les méthodes de terrain	231 - 267
14 / Impliquer la population affectée dans votre évaluation	268 - 282
15 / L'accès restreint	283 - 292
16 / L'analyse	293 - 321

## Notes

# Rédiger et communiquer des résultats d'évaluation

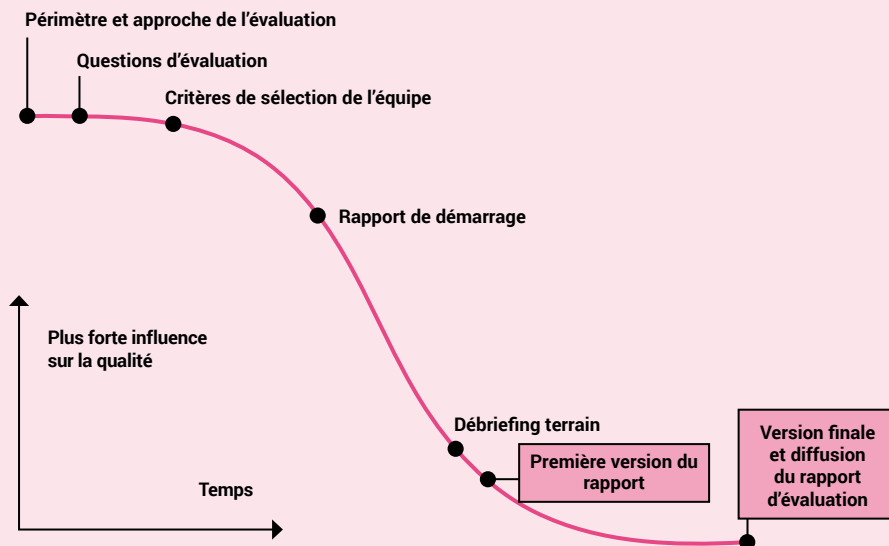


# 17 / Rédiger et communiquer des résultats d'évaluation en se focalisant sur l'utilisation

Comme souligné dans la Section 3 : Penser tôt et souvent à l'utilisation de l'évaluation, il est important d'identifier lors de la planification d'une évaluation ce que les principaux utilisateurs visés doivent savoir et comment cette information peut être communiquée au mieux pour faciliter son utilisation. Comment les résultats de l'évaluation doivent-ils être présentés aux utilisateurs ? En plus du rapport d'évaluation, quels autres produits doivent être disponibles ?

Cette section fournit des éclairages pour que l'équipe d'évaluation rédige ses résultats dans un rapport de haute qualité, attractif et accessible. Il propose également des idées de présentation des résultats de l'évaluation pour maximiser leur utilisation. Enfin, il donne des conseils au responsable d'évaluation et à l'organisation commanditaire au niveau du processus d'approbation, cela afin de garantir un rapport d'évaluation de qualité, ainsi que des activités de diffusion et de communication une fois le rapport terminé.

**Point de contrôle de la qualité :** Première version du rapport et version finale et diffusion du rapport d'évaluation



Cette section couvre deux points clé de la gestion de la qualité. Tout d'abord, la première version du rapport permet au responsable de l'évaluation de se faire une idée de la qualité de l'analyse et d'influer sur la manière dont les résultats sont présentés dans le rapport final. En second lieu, une fois le rapport d'évaluation entre les mains, le responsable de l'évaluation peut se concentrer sur la diffusion des conclusions et des résultats, en ajustant la manière dont ils sont présentés pour s'adapter aux besoins des différents acteurs.

## 17.1 Les produits clés

Les produits demandés à l'équipe d'évaluation comprennent en général : le rapport de démarrage, les ateliers de débriefing, les conseils directement fournis par l'équipe d'évaluation sur le terrain, et le rapport d'évaluation. Le [tableau 17.1](#) résume les principaux produits d'une évaluation, leur timing, et leurs processus d'approbation. Il arrive que l'on demande également à l'équipe d'évaluation de réaliser des activités de diffusion.



Tableau 17.1 : Les produits courants de l'évaluation

Produit	Timing	Processus de révision et de validation
<u>Rapport de démarrage</u>	Au minimum deux semaines avant le travail de terrain pour laisser du temps aux commentaires et aux modifications concernant les méthodes et les calendriers du travail de terrain	Circulation au sein des parties prenantes pour commentaires, en général les parties prenantes ayant un intérêt direct et/ou les principaux utilisateurs visés  Réunion entre le groupe de référence ou de pilotage et l'équipe d'évaluation
Conseils dispensés directement sur le terrain	A mesure que les sujets apparaissent sur le terrain et en réponse à des observations, en particulier lors d'ETR	Aucun
Ateliers de débriefing	A mi-parcours de l'ETR pour les membres de l'équipe d'évaluation dans le but de tester les résultats émergents  A la fin du travail de terrain pour tous les autres types d'évaluation, généralement pour présenter et tester les résultats préliminaires	Les parties prenantes commentent les premiers résultats et les conclusions des évaluateurs en les validant ou en les questionnant.
Note de débriefing	A la fin du travail de terrain (peut être aussi simple que la présentation donnée durant l'atelier de débriefing)	La version finale doit refléter les commentaires reçus lors de la réunion de débriefing autant que les premiers résultats.
Première version du rapport d'évaluation	Trois à quatre semaines après la fin du travail de terrain, un peu plus long pour les évaluations plus complexes (possibilité de plusieurs allers-retours)	Vérification initiale de la qualité par le responsable d'évaluation  Circulation au sein du groupe de référence ou du groupe de pilotage, dont les parties prenantes ayant un intérêt direct et les utilisateurs visés de l'évaluation, pour recueillir leurs commentaires  Réunion ou atelier entre le groupe de référence ou de pilotage et l'équipe d'évaluation pour discuter des points clés
Version finale du rapport d'évaluation	Environ deux semaines après la réception des commentaires finaux sur l'avant-dernière version	Doit être approuvée par le responsable de l'évaluation et/ou un groupe de gestion plus large.

## 17.2 Le rapport d'évaluation

C'est le principal moyen de présentation des résultats, des conclusions et des recommandations de l'évaluation. Comme indiqué dans le *How-To Note on Preparing Evaluation Reports* (USAID, 2012 : 1) :

Un facteur clé de l'utilisation des résultats de l'évaluation réside dans le fait d'avoir un rapport bien écrit et succinct qui communique clairement et rapidement des résultats et des conclusions crédibles, notamment des diagrammes faciles à comprendre et une uniformité de la présentation.

La plupart des organisations ont leurs propres exigences au niveau de la structuration des rapports. Si une certaine flexibilité est possible dans la manière dont le rapport est structuré et que l'organisation commanditaire n'a pas de format défini, l'équipe d'évaluation doit être attentive à ce qui sera le plus utile aux principaux utilisateurs visés de l'évaluation. Structurer le rapport autour des critères d'évaluation peut être utile pour le responsable d'évaluation et pour le département d'évaluation pour qu'ils puissent faire des comparaisons entre évaluations, mais se révéler moins utiles pour les responsables de programme à qui peuvent être adressées de nombreuses recommandations.

A titre d'exemple, si vous évaluez un programme humanitaire multisectoriel, vous souhaitez peut-être structurer au moins une partie du rapport par secteur, par exemple WASH ou abri.

Une alternative consiste à structurer le rapport en fonction des questions d'évaluation. Les évaluateurs doivent envisager la possibilité de partager tôt dans le processus le rapport avec l'organisation commanditaire, par exemple le rapport de démarrage, de manière à garantir qu'il répondra aux besoins des utilisateurs.



### Exemples de structuration de rapport d'évaluation

Pour un rapport structuré d'après les critères OCDE-CAC, voir l'évaluation FICR de ses programmes « moyens de subsistance » suite au séisme en Haïti (Thizy, 2013).

Pour un rapport structuré par grand thèmes étudiés durant l'évaluation, voir l'évaluation conjointe du Cluster Global Logistics (Majewski *et al.*, 2012).

Le document *ALNAP Quality Proforma* (ALNAP, 2005), conçu comme une façon d'évaluer la qualité des rapports d'évaluation humanitaire, constitue un guide utile pour rédiger un rapport d'EAH. Il s'intéresse aux points suivants :

- Les éléments de preuve selon lesquels l'évaluation a bien examiné l'intervention en s'appuyant sur les standards internationaux appropriés (par exemple, le droit international humanitaire et les droits de l'Homme ; le Code de Conduite de la Croix-Rouge/ONG, Sphère).
- Le fait que le rapport s'appuie sur une analyse contextuelle de la zone affectée et de la population (notamment les facteurs pertinents sur le plan historique, social, économique, politique et culturel), qu'il soit mis à profit pour soutenir l'analyse de l'intervention et qu'il contienne une analyse claire de la crise, notamment des événements clés (et une chronologie le cas échéant).
- Le fait que certains sujets transversaux sont traités, notamment l'égalité hommes-femmes, le plaidoyer ou encore la prise en compte des groupes vulnérables et marginalisés.



**Astuce :**

Préparez le plan de votre rapport avant que le travail de terrain ne commence pour vous aider à structurer la collecte et l'analyse des données de manière à ce qu'elles répondent aux besoins du rapport final, réduisant ainsi le travail qu'il demande. C'est particulièrement utile pour rassembler et analyser les données qualitatives collectées lors des entretiens.

Un rapport concis est bien plus susceptible d'être lu qu'un rapport long, mais il doit néanmoins fournir suffisamment d'éléments de preuve pour les résultats, les conclusions et les recommandations. Moins il y a de questions d'évaluation dans les TdR, plus l'équilibre entre ces deux exigences sera facile à trouver.<sup>1</sup>



**Astuce :**




S'il est attendu que le rapport d'évaluation corresponde à un format particulier, cela doit être notifié dans les TdR. La longueur peut être mentionnée mais il peut également s'agir de détails de présentation comme la police de caractère, les tailles de marge, etc. certaines organisations peuvent fournir un modèle de rapport.

## La structure du rapport

Tableau 17.2 : La structure du rapport d'évaluation

Section	Contenu obligatoire
Pages de titre et pages d'ouverture	<p>Informations clés, dont :</p> <ul style="list-style-type: none"> <li>• Le nom de l'intervention évaluée</li> <li>• Le calendrier de l'évaluation et la date du rapport</li> <li>• Les zones (pays, région, etc.) couvertes par l'évaluation – des cartes peuvent être insérées</li> <li>• Les noms des évaluateurs et/ou les affiliations des organisations</li> <li>• Nom de l'organisation qui a commandité l'évaluation</li> <li>• Table des matières incluant également les listes des tableaux, diagrammes, schémas et annexes</li> <li>• Liste des acronymes</li> </ul>
Résumé exécutif	<p>Cette section autonome (généralement, une à trois pages) inclut :</p> <ul style="list-style-type: none"> <li>• Présentation de l'action humanitaire évaluée</li> <li>• Objectifs et public cible de l'évaluation</li> <li>• Les méthodes d'évaluation</li> <li>• Les principaux résultats et conclusions, présentés dans le même ordre que dans le rapport principal</li> <li>• Les principales recommandations</li> </ul>
1. Introduction	<p>Le périmètre et l'objectif de l'évaluation, le public cible, la composition de l'équipe et la structure du rapport</p> <p>Les principales questions d'évaluation</p> <p>Des changements ont-ils eu lieu au niveau des questions d'évaluation proposées dans les TdR?</p> <p>Comment la définition du champ de l'évaluation s'est-elle déroulée ?</p>
2. Méthodes	<p>Une description des principales méthodes utilisées, de leur caractère approprié et de la raison pour laquelle elles ont été choisies.</p> <p>Si vous avez réalisé votre évaluation par rapport à une théorie du changement, vous pouvez l'inclure ici ou le mentionner et l'inclure en annexe.</p> <p>La nature et le périmètre de l'implication de la population affectée</p> <p>Les contraintes clés au niveau de la mise en œuvre de l'évaluation (par exemple : le manque de temps, l'accès restreint à la population affectée, le manque de données au niveau des seuils de référence), et leur impact</p> <p>Tous les biais du processus d'évaluation ou de l'équipe d'évaluation, et comment ils ont été atténués</p> <p>Les standards internationaux utilisés comme points de référence par l'évaluation, par exemple Sphère, LEGS et tout autre cadre conceptuel utilisé, par exemple le cadre relatif à la malnutrition ou les cadres relatifs aux moyens de subsistance mentionnés dans la Section 5 : Elaborer votre évaluation.</p>
3. Contexte	<p>L'analyse contextuelle de la crise à laquelle l'intervention répond, par exemple la zone et la population affectée, les événements clés</p>



Section	Contenu obligatoire
4. Principales sections	<p>Organisés par critères d'évaluation, par questions d'évaluation ou selon un autre cadre qui convient à l'évaluation et à ses utilisateurs visés, ces chapitres présentent les éléments de preuve et les résultats.</p> <div data-bbox="422 264 1070 413" style="background-color: #f0f0f0; padding: 10px; border: 1px solid #ccc;">  <p><b>Astuce :</b> Le fait de présenter les conclusions et les recommandations à la fin de chaque section souligne (et aide à garantir) qu'elles sont bien ancrées dans des éléments de preuve, résultats et conclusions pertinents.</p> </div>
5. Conclusions	<p>Découlent de façon logique des résultats centraux du rapport et les reflètent</p> <p>Offrent une base claire et défendable pour des jugements de a</p> <p>Fournissent des éclaircissements pertinents sur l'intervention qui a été évaluée et sur l'objectif de l'évaluation</p>
6. Recommandations (plus d'éléments sur la rédaction des recommandations à la <a href="#">Section 16</a> )	<p>Doivent être :</p> <p>(a) Claires, pertinentes et applicables, reflétant toutes les contraintes liées au suivi</p> <p>(b) La suite logique des principales conclusions et refléter la consultation des parties prenantes clés</p> <p>(c) Présentées par ordre de priorité, avec un calendrier de mise en œuvre, tout en indiquant qui a la responsabilité du suivi</p> <p>(d) Limitées en nombre</p>
Annexes	<p>Les annexes comprennent en général :</p> <ul style="list-style-type: none"> <li>• Les TdR</li> <li>• La liste des personnes interrogées et des sites visités</li> <li>• La liste des documents consultés et des données secondaires utilisées</li> <li>• Plus de détails sur les méthodes, comme les outils de collecte des données</li> <li>• Les données biographiques des évaluateurs et/ou la justification de la composition de l'équipe</li> <li>• Matrice d'évaluation</li> <li>• La chronologie de la problématique ou de l'action évaluée : c'est essentiel pour toutes les évaluations où le respect des délais est un critère</li> </ul> <p>D'autres annexes peuvent aborder des sujets sur lesquels un développement plus approfondi ne serait pas à sa place dans le rapport principal, ou présenter des résultats issus de méthodes particulières (comme un résumé des réponses à une enquête en ligne).</p> <div data-bbox="422 1182 1070 1294" style="background-color: #f0f0f0; padding: 10px; border: 1px solid #ccc;">  <p><b>Astuce :</b> Utilisez les annexes pour appuyer des éléments et des détails qui encombreraient le rapport principal.</p> </div> <div data-bbox="422 1310 1070 1465" style="background-color: #f0f0f0; padding: 10px; border: 1px solid #ccc;">  <p><b>Astuce :</b> Si le rapport circule principalement sous forme électronique, envisagez de présenter les annexes comme un document séparé pour raccourcir le rapport et le rendre moins intimidant lorsqu'on l'ouvre pour la première fois.</p> </div>

Source : D'après UNEG (2010) et ALNAP (2005)

## L'importance du résumé exécutif

Les organisations peuvent avoir des standards en ce qui concerne la longueur du résumé exécutif, mais on recommande en général deux à cinq pages. Gardez à l'esprit que de nombreuses parties prenantes peuvent avoir uniquement le temps (ou prendre le temps) de lire le résumé exécutif : il ne doit donc pas seulement être une représentation fidèle de votre évaluation mais également se suffire à lui-même. Il peut être utile d'inclure un résumé des principales recommandations, par exemple dans un tableau indiquant « crucial/important/souhaitable » à la fin du résumé exécutif.

Le résumé exécutif ne doit pas figurer dans les premières versions du rapport de façon à éviter de générer des commentaires sur lui plutôt que sur le rapport.



### Astuce :

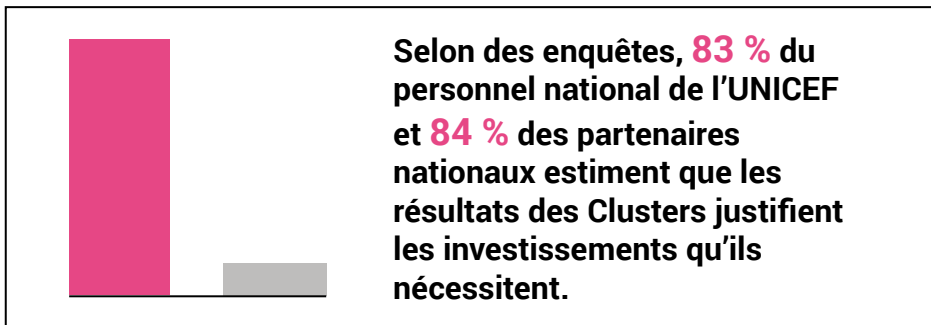
Prévoyez du temps pour rédiger et modifier le résumé exécutif. Ce sera la partie la plus lue du rapport : ne l'écrivez pas à la dernière minute lorsque l'équipe est épuisée !

## Rendre votre rapport plus accrocheur

Il existe de nombreux moyens pour rendre votre rapport plus accessible, lisible et accrocheur. Vous trouverez quelques exemples ci-dessous. Certains sont très faciles à mettre en œuvre, d'autres demandent plus de planification. Ils devront probablement être réfléchis et approuvés avant le travail de terrain.

- **Sous-titres** : Utilisez des sous-titres clairs qui transmettent le principal message de la section ou de la sous-section.
- **Gras** : Le gras (ou l'italique) peuvent aider à faire ressortir des mots clés dans des sections importantes (mais longues) du rapport, par exemple, un mot-clé ou une phrase-clé dans chaque paragraphe. Il est important de rester cohérent dans le type de mots ou de phrases qui sont mis en avant (par exemple : le thème du résultat) (DFID, 2005: 43).
- **Citations en exergue** : Mettez des citations en exergue de manière à faire ressortir toute affirmation particulièrement forte ou importante.
- **Citations directes** : Les citations directes qui sont liées à des résultats clés peuvent aider à donner de la vie au rapport. Même s'il est souvent vital de taire le nom d'une personne, vous pouvez attribuer les citations en utilisant des titres comme « informateur clé de l'organisation partenaire », « instituteur local », etc. (Sutter *et al.*, 2011). Voir la [section 14](#) sur l'obtention du consentement et le respect de la confidentialité.

- **Encadrés** : Des éléments comme les définitions, les mini-études de cas, ou les exemples de bonnes pratiques peuvent être placés dans des encadrés de texte. Ceux-ci peuvent aider à aérer le rapport ainsi qu'à faire ressortir les messages clés ou des histoires.
- **Photographies** : Elles doivent être choisies avec beaucoup de soin et une autorisation de publication adaptée doit être obtenue (mention du droit d'auteur ou autorisation des individus identifiables qui figurent sur l'image). Voir les discussions concernant le consentement éclairé et la confidentialité dans la [Section 14 : Impliquer la population affectée dans votre évaluation](#). Est-ce que l'organisation commanditaire dispose de lignes directrices relatives aux images qui peuvent ou non être utilisées ?
- **Histoires** : Des histoires courtes peuvent être utilisées pour personifier un résultat et aider à rendre un rapport d'évaluation un peu plus vivant. Elles doivent être utilisées de façon modérée pour soutenir ou illustrer un point particulier, puisqu'elles tendent également à trop simplifier les enjeux. Il est important d'obtenir l'autorisation de tout individu évoqué dans l'histoire, en particulier si elle est susceptible d'être utilisée en dehors du rapport ou si le rapport est publié.
- **Tableaux, diagrammes et schémas** : Tableaux, diagrammes, graphiques et autres schémas peuvent communiquer des résultats clés de manière concise. Tufté (1990 ; 1997 ; 2001) propose de bons conseils en matière de présentation des informations visuelles. Vous pouvez aussi trouver quelques exemples dans la [section 10](#).
- **Visualisation des données** : Dans ce domaine, l'étendue des possibilités est en constante évolution, depuis les infographies inanimées jusqu'aux visuels en ligne que le lecteur peut explorer. Pour plus de conseils et d'astuces sur la visualisation des données, voir Evergreen Data (<http://stephanieevergreen.com/blog>).



UNICEF (2013) : Evaluation du rôle d'organisation chef de file Cluster de l'UNICEF pour l'action humanitaire (CLARE), Evaluation Brief



## Différents rapports pour différents publics

Les rapports d'évaluation ont souvent de multiples publics visés. Le responsable de l'évaluation et le personnel du programme qui a mis en œuvre l'intervention évaluée veulent généralement voir l'analyse détaillée, et en particulier les éléments de preuve sur lesquels se fondent les résultats et les conclusions. Ils sont susceptibles de vouloir un rapport détaillé qui expose tout cela. En revanche, d'autres utilisateurs – par exemple, les hauts dirigeants – lisent rarement un long rapport et doivent pouvoir lire un résumé des résultats clés et de leurs implications dans un document concis. Par conséquent, il peut être pertinent de produire différentes versions du rapport final : une version détaillée qui comprend toutes les informations et une version courte autonome. DANIDA publie des notes d'informations pour chacune de ces évaluations. Voir les canaux de communication ci-dessous.

### Exemple de bonnes pratiques : Faire circuler une note d'information

Un résumé autonome de seize pages a été préparé et diffusé pour l'évaluation conjointe du soutien aux PDI, commanditée par Sida, les ministères danois et hollandais des Affaires étrangères, et ECHO en 2005<sup>2</sup>. Ce document rendait beaucoup plus accessibles les résultats clés du rapport de 140 pages.

Voir Borton *et al.* (2005)

## Faire circuler et commenter le rapport provisoire

Les versions provisoires d'un rapport d'évaluation circulent généralement entre le responsable de l'évaluation et les parties prenantes, notamment celles qui ont un intérêt direct et les utilisateurs visés par l'évaluation, la plupart (ou tous) ayant été interrogées durant l'évaluation. Leur donner l'opportunité de répondre à toute erreur de fait, de compréhension ou d'analyse est une façon efficace de garantir la qualité. C'est également une part importante du travail de promotion en faveur de l'appropriation de l'évaluation par les parties prenantes. A l'occasion, il est possible de donner à toutes les personnes interrogées l'occasion de lire et de commenter la version provisoire du rapport, même si cela inclut rarement la population affectée.



## Astuces pour le processus

Lorsque vous envoyez la version provisoire du rapport en demandant aux lecteurs de le commenter :

- Prévoyez suffisamment de temps pour cela : d'une semaine, au minimum, à trois semaines. Il se peut que les destinataires du rapport doivent le faire circuler au sein de leur organisation et rassemblent tous les commentaires.
- Expliquez les limites de temps aux commentateurs. Vous pouvez donner une date limite de remise des commentaires au-delà de laquelle le silence a valeur de consentement.
- Il est utile de signifier aux parties prenantes le niveau de "maturité" du rapport.  
Le niveau ou la profondeur des commentaires qui s'avèrent appropriés à une première version ne sont pas du tout les mêmes que pour une version finale.
- Clarifiez le type de commentaires demandés et comment ils seront partagés. Voir [Exemple de bonnes pratiques à la page 338](#).
- Si quelques personnes seulement commentent le rapport, il est possible de le faire circuler sous format Microsoft Word de manière à ce qu'elles puissent utiliser les fonctions commentaires et suivi de modification, et que chaque commentateur soit identifié. Si de nombreuses parties prenantes commentent le rapport, envisagez de le faire circuler sous format PDF non modifiable, accompagné d'un document Word conçu pour les commentaires.
- Lorsque cela est possible, le responsable d'évaluation doit rassembler tous les commentaires dans un seul document de manière à faciliter leur traitement par l'équipe d'évaluation. Cela aide également à garantir que tous les commentaires relatifs à une section précise sont pris en compte en même temps, rendant ainsi le processus de révision plus efficient.

Une bonne pratique consiste à ce que l'équipe d'évaluation indique comment elle a répondu aux commentaires en fournissant des indications concises pour les commentaires compilés, comme :

- Corrigé
- Nuancé
- Détail ajouté
- Supprimé
- Non accepté – avec une raison détaillée.

**Astuce :**

Numérotez les paragraphes du rapport provisoire afin de faciliter l'écriture des commentaires qui se réfèrent à des passages spécifiques. Ne numérotez pas les lignes car cela rend le processus de révision plus intimidant et il se peut alors que vous vous retrouviez avec plus de commentaires que l'évaluation ne peut en traiter !

**Exemple de bonnes pratiques : La gestion des commentaires à la Croix-Rouge canadienne**

La Croix-Rouge canadienne partage des rapports d'évaluation en interne pour obtenir des retours et vérifier des faits avant de les partager plus largement avec d'autres parties prenantes. Dans le cas de l'évaluation du déploiement de l'Unité de réponse des secours d'urgence aux Philippines après le typhon Yolanda, la version provisoire du rapport d'évaluation a été partagée afin que toutes les personnes interrogées ou ayant répondu à une enquête durant la phase initiale puissent la commenter.

Les commentaires ont été assez utiles et les parties prenantes ont bien accueilli le fait de pouvoir vérifier les résultats. Cela a donné un véritable élan au processus de commentaires et permis que le rapport soit publié plus tôt que prévu quand l'intérêt qu'il suscitait était encore élevé.

Pour rassembler les commentaires, la responsable de l'évaluation a demandé à ce qu'ils soient organisés en trois catégories :

- Corrections et erreurs : toute modification factuelle
- Différences d'opinion ou information complémentaire
- Recommandations : remarques sur ce qui avait été proposé

Elle a proposé aux parties prenantes d'envoyer leurs commentaires par e-mail ou de leur téléphoner pour discuter de leurs idées.

**Valider et finaliser le rapport d'évaluation**

Une fois que l'équipe d'évaluation a répondu à tous les commentaires reçus, elle doit remettre une version finale du rapport pour validation, en général auprès du responsable de l'évaluation ou d'un groupe de gestion. Il existe de nombreux documents de référence très utiles qui peuvent être utilisés par l'organisation commanditaire pour la vérification finale de la qualité, comme le *ALNAP Quality Proforma*, conçu spécifiquement pour l'EAH (ALNAP, 2005), et le *Guide* d'UNEG qui propose une « liste de vérification de la qualité pour les rapports d'évaluation » (UNEG, 2010).

Globalement, le rapport d'évaluation doit être évalué à partir des éléments suivants :<sup>3</sup>

- Répond-il de façon appropriée aux questions d'évaluation ? Si ce n'est pas le cas, les raisons qui expliquent cela sont-elles clairement exprimées et acceptables ?
- Fournit-il des éléments de preuve pour étayer ses résultats ?
- Les conclusions découlent-elles logiquement des principaux résultats du rapport et les reflètent-elles ?
- Les recommandations sont-elles pertinentes par rapport à l'objectif de l'évaluation, et sont-elles étayées par des éléments de preuve et des conclusions ? Voir [Section 16 : L'analyse](#) pour plus de détails sur la manière de passer des éléments de preuve aux recommandations.
- Le rapport est-il cohérent et exempt de contradictions internes ?
- Est-il accessible aux lecteurs et utilisateurs visés, par exemple au niveau de la langue, de sa brièveté et de la clarté de son expression ?
- Respecte-t-il les exigences en matière de présentation et de mise en page ?

## 17.3 La diffusion<sup>4</sup>

### Planifier tôt la diffusion

Bien trop souvent, la planification de la diffusion des résultats et des recommandations de l'évaluation n'a lieu qu'après la finalisation du rapport d'évaluation. Cependant, si vous visez une utilisation sérieuse, c'est l'ensemble de l'évaluation – y compris la diffusion – qui doit être planifiée et budgétée en gardant cet objectif à l'esprit (O'Neil, 2012). Cela va bien au-delà de la simple publication du rapport.

Comme l'expliquent Bamberger *et al.* (2012 : 166) :

De nombreuses évaluations potentiellement utiles ont peu d'impact parce que les conclusions ne sont pas communiquées aux utilisateurs potentiels d'une manière qui leur paraissent utiles ou compréhensibles – ou, pire encore, parce que les conclusions n'ont même jamais atteint des parts importantes de la communauté des utilisateurs.



#### **Exemple de bonnes pratiques : Clarifier la liste de diffusion dans les TdR**

Le « Cadre d'évaluation » de la FICR mentionne que les TdR doivent comprendre une liste de diffusion initiale « pour garantir que le rapport ou la synthèse de l'évaluation atteint le public cible. (...) La diffusion du rapport d'évaluation peut être assurée par différents moyens en fonction du public visé » (FICR, 2011 a : 15).

Voici quelques questions pour vous aider à réfléchir tout au long de la diffusion :

- A quels groupes clés les conclusions et les recommandations de l'évaluation doivent-elles être communiquées ?
- Pourquoi ces informations doivent-elles être communiquées ?
- Que doivent savoir ces différents publics ? Qu'*aimeraient*-ils savoir ?
- Quelle est la meilleure manière de communiquer avec chacun de ces groupes ?
- Faut-il garder à l'esprit certaines considérations ou contraintes particulières propres aux utilisateurs de l'évaluation (par exemple, connexion Internet instable, langue, fort renouvellement du personnel) ?
- Quel est le meilleur timing pour la diffusion (par exemple, la prochaine révision de la stratégie, le nouveau cycle de planification) ?
- Qui a la charge de cette communication ?

## Les rapports d'évaluation dans le domaine public ?

Quelques organisations, en général les plus grandes comme les agences des Nations unies (par exemple, le HCR) et les bailleurs de fonds, ont des politiques claires quant à la diffusion de tous leurs rapports d'évaluation dans le domaine public dans un souci de redevabilité publique. Pourtant, de nombreuses organisations ne le font pas, préférant décider au cas par cas si elles doivent rendre publics leurs rapports. Cela peut augmenter le risque que les rapports d'évaluation plus critiques soient "enterrés", mais dans certains cas les rapports peuvent ne pas être publiés pour des raisons de sécurité, par exemple si le personnel ou les partenaires peuvent être mis en danger à cause de la publication des conclusions.

Une bonne pratique consiste à clarifier dès le début du processus d'évaluation si le rapport sera public ou non. Si nécessaire, les évaluateurs doivent chercher à clarifier ce point lorsqu'ils sont choisis.



### Astuce :

Les interactions personnelles sur le terrain, lors de briefings et d'ateliers, sont plus susceptibles de générer de l'apprentissage que les rapports écrits. Assurez-vous d'en planifier, en particulier pour les évaluations qui se concentrent sur l'apprentissage.

## Concevoir une stratégie de diffusion

En tant que responsable d'évaluation, vous devez réfléchir à la conception d'une stratégie de diffusion pour chacune de vos évaluations, quel que soit l'étendue ou le niveau de la communication et de l'adoption des conclusions de l'évaluation. Une stratégie de diffusion signifie l'identification des canaux de communication et des produits qui conviennent le mieux aux besoins des différents publics et utilisateurs de l'évaluation. Certaines recommandations peuvent avoir différents publics cibles. Parfois, une stratégie de diffusion doit être conçue pour chaque conclusion et chaque recommandation pour être sûr qu'elles atteignent bien leur public cible et qu'elles sont utilisées de façon adéquate. A titre d'exemple, les recommandations stratégiques et opérationnelles font sens à des niveaux assez différents d'une organisation. Il est possible que vous souhaitiez envisager différentes méthodes de communication pour ces deux publics, chacune se concentrant sur ce que chaque public a besoin de savoir.

## Communiquer et faire des retours aux parties prenantes et aux personnes interrogées sur le terrain

En dehors des ateliers de débriefing de la fin de la phase terrain, on constate souvent peu d'investissement dans la communication des conclusions de l'évaluation aux parties prenantes et personnes interrogées du pays. Si les principaux utilisateurs visés de l'évaluation sont sur place, il faut leur accorder l'attention voulue. Envisagez la possibilité de budgéter le retour des membres de l'équipe d'évaluation sur le(s) site(s) de l'évaluation pour des compte-rendus aux parties prenantes clés, comme le personnel du programme. Envisagez également de commenter les conclusions avec les membres de la population affectée qui ont été impliquées dans l'évaluation. Voir l'exemple de bonnes pratiques du WVI ci-dessous.



### **Exemple de bonnes pratiques : World Vision International fait des retours aux populations affectées**

World Vision International utilise des processus de validation communautaire pour faire des retours aux communautés qu'elle a consulté, soit durant les diagnostics, soit durant les consultations communautaires, soit durant les évaluations. Ces processus sont utilisés pour informer les communautés quant à la manière dont les consultations ont influencé les décisions de l'organisation. Le personnel de WVI staff peut réaliser des affiches, utiliser des *paper-boards* et d'autres supports pour faciliter la présentation des conclusions auprès des enfants et des adultes. Certaines méthodes plus créatives ont eu recours à des marionnettes pour les enfants.

## 17.4 Les différents canaux de communication

Cette section présente un certain nombre de canaux et de produits différents pour partager les résultats ou leçons de l'évaluation, que ce soit au siège, dans le pays de l'évaluation, ou de façon plus large au sein d'une organisation. La liste qui suit est loin d'être exhaustive : soyez créatif.

### Réunions, communication personnelle et briefings

- Des réunions ciblées peuvent être organisées entre l'équipe d'évaluation et les utilisateurs visés. Si elles sont conçues sous la forme d'ateliers et adaptées aux besoins de chaque groupe, elles peuvent encourager une plus grande implication des utilisateurs.



#### Astuce :

Les présentations synthétiques utilisées pour les réunions et les ateliers peuvent-elles circuler de façon large ? Peuvent-elles être traduites ? L'organisation commanditaire pourrait ainsi faire circuler plus facilement les conclusions de l'évaluation sur le plan local.

- Des briefings en tête-à-tête peuvent être proposés aux utilisateurs clés par l'équipe d'évaluation ou le responsable de l'évaluation.
- Rédiger différentes versions du rapport ainsi que des résumés très concis et accessibles ou encore des notes de briefing. Comme l'écrit Chris Lysy dans son blog BD, FreshSpectrum :<sup>5</sup> « N'écrivez pas un seul rapport de 200 pages. Rédigez un rapport de 60 pages, deux rapports de 30 pages, cinq rapports de 10 pages, dix rapports de 5 pages, et vingt rapports d'une page : chacun avec un objectif différent, personnalisé pour chaque type de public » (août 2014).
  - ODI RAPID utilise la règle du 1-4-25 : 1 page pour les décideurs, 4 pages pour informer et 25 pages pour le rapport dans sa forme complète.
  - L'évaluation UNICEF de son rôle de chef de file Cluster pour l'action humanitaire est un bon exemple de document de 4 pages (UNICEF, 2014).
  - Voir les notes de synthèse du Bureau de l'évaluation du PAM.<sup>6</sup>



#### Astuce :

S'il n'est pas prévu que le rapport complet de l'évaluation soit traduit dans les langues pertinentes, envisagez de traduire au moins le résumé exécutif ou la note de briefing dans les langues locales pour en assurer la diffusion aux parties prenantes locales.

- Pensez à ajouter une « page de garde » au rapport d'évaluation qui classe les leçons apprises et donne un aperçu des raisons pour lesquelles ce rapport pourrait intéresser le lecteur (Oliver, 2007: 18).
- Des e-mails brefs peuvent communiquer les résultats, conclusions, ou recommandations aux utilisateurs clés.



**Astuce :**

Quels supports rédigés sur-mesure à partir de l'évaluation pourraient être fournis aux responsables ? Par exemple, est-ce que des points de discussion peuvent être rédigés à l'attention des responsables susceptibles de devoir faire un compte-rendu de l'évaluation durant des réunions réservées à cet effet ?

## Outils audio-visuels

- **Vidéos** : Les conclusions et les recommandations peuvent être communiquées au moyen d'un enregistrement vidéo de l'évaluation ou des entretiens enregistrés avec l'équipe d'évaluation. Cela peut générer des retours très intéressants sur des sujets clés en utilisant les propres mots de la population affectée, retours dont les gens se souviendraient plus longtemps que des rapports d'évaluation. Voir [l'exemple de bonnes pratiques à la page 344](#) et la discussion sur le [consentement éclairé dans la Section 14](#).



**Astuce :**

Le logiciel PowerPoint permet d'ajouter des enregistrements audio aux diapositives et de convertir la présentation en une vidéo. Si vous ne pouvez pas vous rendre en personne à une réunion, il est possible d'envoyer un résumé-vidéo des conclusions de l'évaluation.

- **Podcasts** (enregistrements audio pouvant être écoutés sur un lecteur MP3 et un ordinateur) discutant des points clés de l'évaluation.
- **Webinar** (discussions en panel ou présentations, en ligne), comme le webinar ALNAP co-hébergé par DEC qui traite de l'utilisation de la contribution au changement aux Philippines ([www.alnap.org/webinar/19](http://www.alnap.org/webinar/19)) ou les séries webinar d'Interaction sur les évaluations d'impact ([www.interaction.org/impact-evaluation-notes](http://www.interaction.org/impact-evaluation-notes)).



### Exemple de bonnes pratiques : Utiliser la vidéo pour diffuser des messages ou des leçons clés

Le Bureau de l'apprentissage, de l'évaluation et de la recherche d'USAID a engagé un vidéaste pour accompagner une équipe d'évaluation sur le terrain. Ce documentariste a suivi une équipe d'évaluation de projet durant trois semaines dans le but de documenter la plupart des phases de l'évaluation. Un des objectifs était de générer de l'apprentissage sur le processus d'évaluation, de montrer certains des défis rencontrés sur le terrain (en particulier dans les zones de conflit), et de permettre aux personnes qui gèrent, commanditent, conçoivent et réalisent des évaluations de mieux comprendre la complexité et la valeur de l'évaluation. La vidéo est disponible à l'adresse suivante : [www.usaidlearninglab.org/library/evaluating-growth-equity-mindanao-3-program](http://www.usaidlearninglab.org/library/evaluating-growth-equity-mindanao-3-program)

La vidéo du Groupe URD sur l'ETR Haïti peut être vue en ligne disponible à l'adresse suivante : [vimeo.com/15198053](https://vimeo.com/15198053). La vidéo d'OCHA sur l'ETR inter-organisations des inondations au Pakistan (2011) est visible ici : [ocha.smugmug.com/Film/OCHA-Films/i-rdN3hsX/A](http://ocha.smugmug.com/Film/OCHA-Films/i-rdN3hsX/A)

Source : Paul Barese dans AEA 365 [www.aea365.org/blog/?p=8768](http://www.aea365.org/blog/?p=8768) Linda Morra Imas, IDEAS Evaluation listserv exchange, avril 2013

### Les solutions via Internet : réseaux sociaux et blogs

Les réseaux sociaux évoluent rapidement, et de nouvelles plateformes et utilisations émergent continuellement. Les réseaux sociaux sont de plus en plus utilisés pour annoncer la sortie des rapports d'évaluation ou des conclusions et recommandations de l'évaluation. A l'heure actuelle, la plupart des plateformes des réseaux sociaux s'orientent vers un public externe. Elles offrent néanmoins beaucoup de place à la créativité.

L'une des règles clés des réseaux sociaux est d'apprendre leurs règles de fonctionnement ou leurs fonctionnalités spécifiques, et d'ajuster le contenu en fonction. Par exemple, il peut être pertinent d'annoncer la publication d'un nouveau rapport d'évaluation via Twitter, mais cela peut donner lieu à un post de blog ou de forum un peu plat, pour lesquels des leçons apprises pourraient être plus intéressantes.

Voici quelques exemples :

- **Blogs** : Les détails relatifs aux avancées de l'évaluation de la transversalisation Age-Genre-Diversité du HCR pour la Colombie (Mendoza et Thomas, 2009) ont été mis à jour sur un blog intitulé « It Begins with Me. It Begins with You. It Begins with Us ». (Voir [itbeginswithme.wordpress.com](http://itbeginswithme.wordpress.com)).
- **Reportage photos** : Les images peuvent être présentées sous format PowerPoint ou PDF mais aussi via des réseaux sociaux comme



Flickr.com, plateforme de partage de photos qui se prête assez bien à cela. Voir l'exemple d'un reportage photo sur l'évaluation par la Banque de développement asiatique : [www.flickr.com/photos/71616117@N06/sets/72157630309259904/with/7452881380](http://www.flickr.com/photos/71616117@N06/sets/72157630309259904/with/7452881380).

- **Forum Communauté de pratique sur l'évaluation humanitaire d'ALNAP** : [partnerplatform.org/alnap/humanitarian-evaluation](http://partnerplatform.org/alnap/humanitarian-evaluation).
- **Groupe LinkedIn** : La Société européenne d'évaluation conduit des discussions sur LinkedIn.
- **Twitter** : Le Groupe URD (@GroupeURD), par exemple, tweete de façon très active sur son travail d'évaluation.
- **Reddit** : Comparable à un tableau d'affichage en ligne, Reddit laisse ses utilisateurs soumettre des liens que d'autres utilisateurs notent, créant ainsi un classement virtuel. Le contenu est organisé par sujets appelés « subreddits ». Il en existe un spécifiquement consacré à l'évaluation ([www.reddit.com/r/Evaluation](http://www.reddit.com/r/Evaluation)).

## Bases de données Evaluations

Si une personne ne peut pas trouver votre rapport, elle ne peut pas l'utiliser. De nombreuses organisations investissent dans l'amélioration de leurs bases de données Evaluations. Par exemple, la FICR a restructuré sa base de données en ligne (voir [www.ifrc.org/fr/publications/evaluations](http://www.ifrc.org/fr/publications/evaluations)), et Tearfund a développé un outil Google Docs pour partager au sein de toute l'organisation des détails sur les évaluations planifiées et terminées (Hallam et Bonino, 2013 ; Warner, 2014).

Créée à la fin des années 90, la Evaluation Library<sup>7</sup> d'ALNAP offre la collection la plus complète de documents évaluatifs du domaine humanitaire à ce jour – des rapports d'évaluation, mais aussi des méthodes d'évaluation et des guides méthodologiques, ainsi qu'une sélection de documents sur la recherche en matière d'évaluation. Assurez-vous de bien mettre en ligne vos produits d'évaluation avec ALNAP<sup>8</sup>



### Astuce :

Incluez les annexes dans la version en ligne du rapport d'évaluation. Il est également possible de mettre en ligne une version sans annexes. Cela évite le cas où une personne sera capable de trouver le rapport dans cinq ans, mais pas les annexes.

## 17.5 Faciliter l'adoption d'une évaluation

Un certain nombre de facteurs influencent l'utilisation des conclusions et des résultats d'une évaluation. Voir [Section 3](#) et Hallam et Bonino (2013), pour plus d'informations à ce sujet. Par exemple, l'évaluation fait-elle l'objet d'une demande en interne ? La renforcer et la maintenir est un processus de plus haut niveau et de plus longue durée auquel une seule évaluation n'apporte qu'une simple contribution.

Comme indiqué précédemment, impliquer les principales parties prenantes tout au long du processus d'évaluation est essentiel pour l'utilisation et l'adoption des conclusions (voir [Section 4](#)), tout comme l'est une phase de diffusion bien planifiée et bien financée (voir [Section 17.4](#)).

L'adoption peut également être facilitée par :

- L'implication de certains des utilisateurs clés dans la formulation des recommandations. Cela peut renforcer l'appropriation des recommandations par les parties prenantes.
- L'identification des personnes influentes au sein de l'organisation qui sont impliquées dans des actions concrètes et dans le changement.
- L'intégration des conclusions de chaque évaluation (ou de plusieurs évaluations portant sur le même sujet) dans des supports de formation, par exemple des études de cas.
- La désignation claire des responsabilités en matière de suivi. Dans les grandes organisations, cela se fait généralement via un tableau de gestion des recommandations.



### Exemple de bonnes pratiques : Développement ou perfectionnement consultatif des recommandations

L'équipe de l'évaluation de Yogyakarta n'a rédigé aucune recommandation provisoire. A la place, les conclusions ont été discutées lors d'un atelier multi-parties prenantes organisé à Yogyakarta, qui comprenait des bénéficiaires, des représentants du gouvernement, des ONG locales et internationales, et du personnel des quatre organisations évaluées. Les parties prenantes ont étudié et modifié les conclusions puis formulé certaines recommandations aux ONG par rapport à de futures interventions. (Wilson *et al.*, 2007 : 2-3)

Dans le cas du Pakistan, l'équipe d'évaluation a proposé aux parties prenantes des recommandations à améliorer. Une fois le rapport provisoire préparé, l'équipe est retournée au Pakistan pour organiser trois ateliers en province et un atelier au niveau national avec les parties prenantes clés impliquées dans la réponse humanitaire aux inondations.



Le chef d'équipe a présenté les résultats, les conclusions et les recommandations durant les ateliers. Par la suite, les parties prenantes ont conjointement validé et priorisé les recommandations puis défini qui serait responsable de la mise en œuvre de chaque recommandation et de son calendrier. Les principaux changements de formulation ont résulté des discussions de groupe. L'équipe d'évaluation a estimé que ce processus a accéléré l'appropriation des recommandations de l'évaluation et renforcé l'apprentissage en temps réel parmi les parties prenantes impliquées. (Polastro *et al.*, 2011 : 4).

World Vision International a également recours aux ateliers pour développer les recommandations de certaines évaluations. Ses équipes rassemblent un groupe important de parties prenantes, parfois jusqu'à 100, regroupés par équipes, dans un atelier d'interprétation qui peut durer jusqu'à deux jours. Les participants peuvent être des hauts dirigeants, du personnel sectoriel, des partenaires du projet, etc. Durant le processus, différentes sections des conclusions de l'évaluation sont fournies à chaque groupe plutôt que le rapport entier. Ces sections peuvent par exemple être organisées autour des critères OCDE-CAD. Ensuite, chaque groupe doit travailler sur sa section du rapport pour répondre à une série de questions. C'est un levier permettant aux participants de comprendre comment interpréter les conclusions pour pouvoir par la suite les utiliser de façon adéquate durant le reste de l'atelier dans le but final de générer des recommandations. De cette manière, les participants prennent confiance dans l'utilisation des conclusions et sont bien plus à même de se référer au rapport dans le futur pour des informations et des prises de décision.

Source : Communication personnelle : K. Duryee, World Vision International, 2013 ; Chamberlain, Jensen et Cascioli Sharp (2014) ; Vallet (2014).



**Définition : Tableau de gestion des recommandations**

Registre de la réponse associée à chaque recommandation de l'évaluation ainsi que des étapes que les dirigeants prévoient de suivre pour y répondre, comprenant une date butoir et la partie responsable de chaque étape.

Schéma 17.3 : Exemple de tableau de gestion des recommandations

	Nécessité d'un financement supplémentaire ?		Réponse de la direction			Commentaire	Action à mener	Timing	Unité responsable
	Oui	Non	Acceptée	Partiellement acceptée	Rejetée				
Recommandation 1									
Recommandation 2									

Source : FAO

L'intérêt du tableau de gestion des recommandations réside dans le fait d'encourager les discussions entre utilisateurs visés sur les recommandations ainsi que sur la manière dont ils les mettront en œuvre (ou non). Il peut également servir d'outil de redevabilité (UNEG, 2010). Certaines organisations, comme le PAM et la FAO, ont un processus de suivi du tableau de gestion des recommandations, six mois ou un an après, pour contrôler que l'action proposée a bien été mise en œuvre.

Les points qui suivent, produits par UNEG (2010 : 7-8), sont des bonnes pratiques en matière de gestion des recommandations. Ils sont pertinents aussi bien pour les grandes évaluations que pour les petites.

- **Des rôles et des responsabilités** clairement définis doivent être communiqués à toutes les parties prenantes de l'évaluation.
- **Définir une date butoir concertée** à laquelle la direction ou d'autres parties prenantes clés doivent fournir leur réponse formelle à l'évaluation.
- La direction doit **nommer un référent** pour coordonner la gestion des recommandations.
- La gestion des recommandations doit clairement indiquer si la **direction accepte, accepte partiellement ou rejette les recommandations**. Dans le dernier cas, la/les raison(s) doivent être données. Dans les deux premiers cas, les actions à entreprendre doivent être mentionnées en détail, indiquant un calendrier précis et la/les unité(s) chargée(s) de les mettre en œuvre. Lorsque plus d'une unité est mentionnée, il faut indiquer clairement qui est responsable de quelle(s) action(s).
- La réponse de la direction des recommandations doit être **communiquée en liaison avec le rapport d'évaluation**.

Le tableau de gestion des recommandations se révèle limité lorsqu'un certain nombre d'organisations ont été impliquées dans une évaluation conjointe. Néanmoins, il est encore possible qu'une organisation identifie les recommandations qui sont pertinentes pour elle et élabore son propre tableau de gestion des recommandations. UNEG (2010), dans ses points signalés comme des bonnes pratiques, suggère de créer un groupe ad hoc de gestionnaires des recommandations au sein des différent(e)s organisations/partenaires impliqué(e)s dans une évaluation conjointe de façon à obtenir un suivi coordonné.



**Exemple de bonnes pratiques : Impliquer les parties prenantes clés tout au long de l'évaluation**

Pour l'évaluation ECB de la réponse au séisme de Yogyakarta (Wilson *et al.*, 2007), l'équipe d'évaluation a présenté un résumé des résultats de l'évaluation au comité de pilotage et au personnel de terrain de manière à ce que ces parties prenantes clés puissent travailler avec l'équipe pour rédiger les conclusions et recommandations. Une réunion finale a été organisée entre l'équipe d'évaluation, les membres du comité de pilotage, des représentants du gouvernement et des habitants de la région pour analyser et modifier les premières conclusions et établir de nouvelles recommandations.

## 17.6 Les synthèses d'évaluation, les revues thématiques et les méta-analyses pour faciliter l'adoption

Comme l'expliquent Hallam et Bonino (2013 : 78-80) : La plupart des utilisateurs potentiels des résultats d'une évaluation veulent en savoir plus que ce qu'une seule évaluation ou étude a trouvé. « Ils veulent connaître le poids des éléments de preuve » (Weiss, 1998 : 317). Des dizaines d'études, d'évaluations et de revues thématiques peuvent couvrir le même sujet ou enjeu. Les observer dans leur totalité, au moyen d'une méta-analyse et d'une synthèse des évaluations, peut produire des éléments de preuve et des résultats bien plus riches (Banque mondiale, 2009 : 70).



**Définition : Synthèse des évaluations**

Une synthèse des évaluations suit une procédure systématique pour organiser et résumer les résultats d'une série d'évaluations (potentiellement assez disparates). Cela peut s'avérer une opportunité pour poser des questions précises sur la série d'évaluations, et/ou pour identifier des tendances courantes dans les résultats.

**Définition : Revue thématique**

Une revue thématique extrait les résultats et les leçons apprises d'une série d'évaluations de l'action humanitaire dans un sous-secteur, par exemple la protection, l'habitat/abri. Ce type d'étude peut être commandité pour répondre à / traiter une série précise de questions fondamentales.

**Définition : Méta-évaluation**

Une méta-évaluation est une évaluation globale conçue pour agréger des résultats issus d'une série d'évaluations, généralement en fonction d'un ensemble de critères. Elle peut également être utilisée pour désigner l'« évaluation d'une évaluation », c'est-à-dire pour juger la qualité d'une série d'évaluations.

Il existe plusieurs guides utiles portant sur les processus de synthèse. Même s'ils ne sont pas tous spécialisés sur l'EAH, leurs principes sont néanmoins valables. Par exemple, le guide d'Oxfam GB et du Feinstein International Center (n.d.) développé pour leur Humanitarian Evidence Programme vise à synthétiser les recherches dans le secteur humanitaire.

Plusieurs organisations rassemblent et synthétisent les leçons apprises au cours d'une série d'évaluations de l'action humanitaire :

- CARE a réalisé une méta-analyse des évaluations et des revues après action à partir de quinze réponses d'urgence, et en a tiré des leçons clés (Oliver, 2007 : Annex 1A).
- Le rapport d'évaluation annuel du Bureau d'évaluation du PAM formule des recommandations stratégiques pour le PAM en s'appuyant sur la synthèse des résultats, conclusions, leçons et recommandations de toutes les évaluations réalisées par le Bureau d'évaluation durant l'année précédente, et rend compte des activités évaluatives associées pour renforcer la fonction d'évaluation au sein du PAM et du système international : [www.wfp.org/content/annual-evaluation-report-2013-0](http://www.wfp.org/content/annual-evaluation-report-2013-0).
- NORAD produit régulièrement un rapport de synthèse des leçons apprises des évaluations (Disch *et al.*, 2008 ; Stokke, 2007).
- ACF et Tearfund produisent des rapports d'apprentissage annuels fortement inspirés des résultats des évaluations. Il s'agit respectivement des *Learning Review* (ACF, 2014) et des *Learning and Impact Report* (Tearfund, 2015).

- Depuis 2013, les agences des Nations unies réalisent des méta-évaluations annuelles pour étudier dans quelle mesure le genre a été intégré de façon adéquate dans les rapports d'évaluation en utilisant un modèle développé par UNEG. Ces méta-évaluations sont utilisées à des fins de redevabilité pour évaluer si les agences répondent aux exigences minimales spécifiées dans les Normes et standards pour l'évaluation d'UNEG. Début 2016, par exemple, le PAM a commandité des consultants indépendants pour une méta-évaluation de ce type. Dans le cas du PAM, la plupart des évaluations relèvent de l'assistance humanitaire.

Certaines méta-revues ont exploré dans quelle mesure les résultats et recommandations des évaluations passées ont été prises en compte (Oliver, 2009). Elles peuvent se concentrer sur des sujets où l'on constate des difficultés d'apprentissage persistantes, et peuvent étudier de façon utile les raisons de la résistance au changement, mais aussi proposer des solutions pour lever de tels blocages à l'avenir.

Les *Lessons Papers* d'ALNAP visent à rendre disponibles les leçons apprises des réponses précédentes sous une forme concise et facile à lire, de manière à informer et ainsi améliorer dans le futur l'efficacité de l'action humanitaire. Ils sont destinés au personnel en charge de concevoir et mettre en œuvre des réponses humanitaires. Ces documents se fondent principalement sur les conclusions d'évaluations continues dans le Humanitarian Evaluation and Learning Portal (HELP) d'ALNAP, ainsi que dans des documents complémentaires. Tous les *Lessons Papers* d'ALNAP sont disponibles à l'adresse suivante : [www.alnap.org/what-we-do/lessons](http://www.alnap.org/what-we-do/lessons).

## Notes de fin de texte

1. Les contenus de rapport décrits dans cette section s'inspirent du document ALNAP Quality Proforma (2005).
2. Voir [www.sida.se/contentassets/ea2b7ee9cdff40a7a0a54022ebf1167d/support-to-internally-displaced-persons-learning-from-evaluation.-synthesis-report-of-a-joint-evaluation-programme—summary-v\\_3325.pdf](http://www.sida.se/contentassets/ea2b7ee9cdff40a7a0a54022ebf1167d/support-to-internally-displaced-persons-learning-from-evaluation.-synthesis-report-of-a-joint-evaluation-programme—summary-v_3325.pdf).
3. D'après UNEG (2010) et ALNAP (2005).
4. Cette section est fortement inspirée de Hallam et Bonino (2013), Capacity Area 3, et Warner (2014).
5. Voir [freshspectrum.com](http://freshspectrum.com).
6. Voir [www.wfp.org/evaluation/lessons/evaluation-briefs](http://www.wfp.org/evaluation/lessons/evaluation-briefs).
7. Voir [www.alnap.org/resources/results.aspx?type=22](http://www.alnap.org/resources/results.aspx?type=22).
8. [www.alnap.org/account/submitresource.aspx](http://www.alnap.org/account/submitresource.aspx).



## Notes

## Notes

# Les évaluations d'impact humanitaire



# 18 / Les évaluations d'impact humanitaire

## 18.1 Pourquoi une évaluation d'impact est-elle importante ?

La plupart des EAH se concentrent sur les résultats, souvent de façon descriptive. En d'autres termes, elles décrivent si un résultat visé a été atteint, par exemple si les taux de malnutrition ont chuté, mais ont du mal à établir un lien de causalité entre le résultat et le programme ou l'intervention évalué(e)<sup>1</sup>. Il est encore moins probable qu'elles saisissent l'impact relatif aux effets les plus larges de l'action humanitaire, même si les TdR de la plupart des évaluations de l'action humanitaire comprennent des questions liées à l'impact (Proudlock et Ramalingam, 2009 ; Knox-Clarke et Darcy, 2014: 40-44). On trouve encore peu d'exemples d'EAH qui se concentrent uniquement sur l'impact. De nombreuses raisons expliquent cela, certaines étant décrites ci-dessous dans la sous-section sur les difficultés.

L'intérêt croissant pour l'évaluation d'impact de l'action humanitaire est dû au fait que les bailleurs et les organisations opérationnelles veulent déterminer l'impact des actions humanitaires financées par des fonds internationaux sur les vies et les moyens de subsistance des personnes qu'elles cherchent à aider. A titre d'exemple, les bailleurs posent des questions sur l'impact de leur investissement – souvent important – dans l'action humanitaire, en général du point de vue de la redevabilité. Il est enfin possible que les organisations opérationnelles veuillent savoir quelles réponses humanitaires sont les plus efficaces, souvent à des fins d'apprentissage.

## 18.2 Qu'est-ce que l'impact dans l'EAH ?

La définition de l'impact, d'après les critères OCDE-CAD, est la suivante : L'impact observe les effets les plus larges du programme – sociaux, économiques, techniques, environnementaux – sur les individus, les groupes définis par le sexe et l'âge, les communautés et les institutions. Les impacts peuvent être intentionnels ou non, positifs et négatifs, macro (sectoriels) et micro (ménages, individus), à court ou long terme. (d'après Beck, 2006: 21)

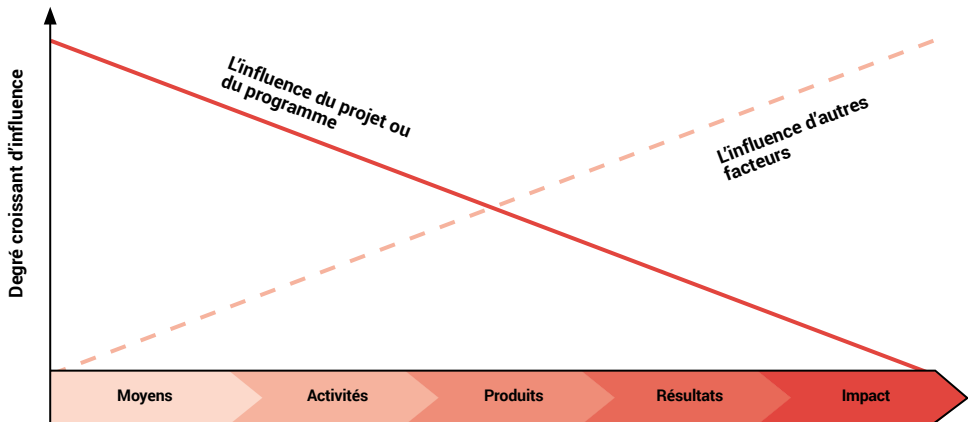
Dans la pratique, les organisations interprètent l'impact de différentes manières par rapport à l'évaluation de l'action humanitaire. Certaines se concentrent sur les *résultats qui peuvent* être attribués à l'intervention. Par exemple, une réduction des

taux de malnutrition peut-elle être attribuée à la distribution d'aide alimentaire ? D'autres organisations veulent savoir des informations sur l'impact le plus large, notamment les effets dominos de l'action humanitaire, intentionnels ou non, positifs et négatifs. Par exemple, dans quelle mesure la distribution de l'aide alimentaire a-t-elle affecté les marchés céréaliers ou la production alimentaire locale ?

L'essence de l'évaluation d'impact est d'explorer les causes et les effets, autrement dit l'inférence causale. Elle éloigne l'analyse de l'efficacité et de l'efficacité de l'intervention, pour observer si les populations vont mieux ou sont plus en sécurité grâce à elle. Déterminer ce lien de causalité est un véritable défi.

Dans certains contextes, une évaluation d'impact peut être demandée pour étudier les changements persistants et significatifs de long-terme qui peuvent être attribués à l'action humanitaire. Après une catastrophe naturelle, comme des inondations ou des sécheresses, on peut ainsi constater des éléments de preuve liés à l'amélioration de la préparation ou de la résilience aux catastrophes à venir. Dans un conflit de longue durée présentant des déplacements massifs et durables, comme au Soudan du Sud ou dans le cadre de la crise syrienne, l'évaluation peut se concentrer davantage sur l'impact immédiat, par exemple l'impact de l'aide humanitaire fournie aux personnes déplacées sur la population hôte, plutôt que sur l'impact dans la durée.

Schéma 18.1 : L'influence relative tout au long de la chaîne des résultats



Source : Smutylo (2001)

Le défi fondamental de l'évaluation d'impact est l'attribution, c'est-à-dire le fait d'isoler l'impact causé par une intervention donnée des nombreux autres facteurs en jeu. Comme démontré dans le [Schéma 18.1](#) (Smutylo, 2001), l'influence de l'action d'une seule organisation (ou même de plusieurs organisations) décroît au fil de la chaîne des résultats à mesure que l'influence d'autres facteurs augmente. Par exemple,

- Même si les organisations humanitaires peuvent avoir distribué efficacement l'aide alimentaire et fourni des soins de santé primaires à la population affectée, l'impact positif de telles interventions peut avoir été minimal si le conflit s'est intensifié, que ces communautés se sont déplacées et que la crise humanitaire s'est aggravée.
- A l'inverse, la sécurité alimentaire d'une population affectée peut s'être améliorée sans que l'on puisse clairement attribuer dans quelle mesure cela est dû à l'action humanitaire. Par exemple, d'autres facteurs extérieurs peuvent avoir joué un rôle, comme lorsqu'une amélioration de la sécurité permet que les marchés fonctionnent de façon plus efficace, ou que des facteurs saisonniers comme la récolte augmentent la quantité de nourriture disponible.

La difficulté d'attribution n'est pas propre à l'EAH, mais elle est amplifiée dans la plupart des contextes humanitaires par la complexité liée au fait de rassembler des preuves de grande qualité. Cela peut s'expliquer par les défis contextuels suivants des crises humanitaires, comme décrit dans la [Section 1 : Qu'est-ce que l'évaluation de l'action humanitaire ?](#) :

1. **Défi** : La dynamique et l'environnement fluctuant de la plupart des contextes humanitaires, dont les nombreux facteurs imprévisibles affectent les résultats et l'impact, notamment une grande pluralité d'acteurs.

**Solution potentielle** : envisagez des évaluations d'impact conjointes (voir ci-dessous)

2. **Défi** : Le manque de données, notamment :
  - a. Les données de base nécessaires pour concevoir certaines méthodes d'évaluation, comme sur la population ou le nombre de personnes affectées par la crise.
  - b. Les données de seuil de référence sur des indicateurs clés liés au bien-être, par exemple sur les moyens de subsistance ou l'accès à l'éducation, à partir desquels il est possible d'évaluer si un changement a eu lieu.
  - c. Les données de suivi disponibles et de grande qualité qui illustrent le changement au fil du temps (les données de suivi se concentrent en général sur le processus et les produits plutôt que sur les résultats)

**Solution potentielle :** Faites appel aux souvenirs, et demandez aux membres de la population affectée quelle était leur situation auparavant et en quoi elle a changé ; utilisez des méthodes alternatives pour compenser l'absence de données de base, par exemple l'imagerie satellitaire pour compter les habitations.

En plus de ces difficultés contextuelles, certaines difficultés propres à l'évaluation affectent l'évaluation dans le secteur humanitaire :

1. Le besoin d'une action rapide dans un environnement imprévisible. Cela signifie qu'une évaluation d'impact ne bénéficie que de peu de temps pour sa préparation, à partir des premières phases de la crise et de la réponse (Puri *et al.*, 2014). Les évaluations d'impact tendent à être planifiées tard dans le cycle du programme.
2. Choisir la conception la plus adéquate et la combinaison d'approches qui convient le mieux pour répondre à des questions spécifiques portant sur les liens de causalité.
3. L'évaluation d'impact demande des niveaux de compétences différents de ceux des évaluations conventionnels. Les exigences liées à la collecte des données et à l'analyse peuvent nécessiter une palette de compétences plus orientées vers la recherche. Ces compétences sont généralement rares dans le secteur humanitaire (Alexander et Bonino, 2015 ; Proudlock *et al.*, 2009).

Du fait de ces difficultés, et tout particulièrement de celle liée au fait d'isoler l'impact d'une intervention particulière, l'évaluation d'impact de l'action humanitaire tend à se concentrer davantage sur l'attribution partielle ou sur la contribution d'une intervention spécifique à un changement (voir [Section 1 : Qu'est-ce que l'évaluation de l'action humanitaire ?](#)).

## 18.3 Planifier une évaluation d'impact de l'action humanitaire

Les difficultés propres à la réalisation d'une évaluation d'impact expliquent que la phase de planification doit bénéficier de beaucoup de temps. Posez-vous les questions suivantes :

### **Les conditions sont-elles réunies pour une évaluation d'impact ?<sup>2</sup>**

Les évaluations d'impact de grande qualité demandant des ressources considérables, assurez-vous qu'au moins une des conditions suivantes est remplie pour justifier l'investissement :

- L'intervention est-elle suffisamment importante (en termes de taille, d'importance politique ou de conséquences potentielles) pour demander un tel type d'évaluation ?

- Est-elle pertinente sur le plan stratégique en termes de leçons apprises et d'influences potentielles d'une telle évaluation d'impact ?
- L'évaluation d'impact proposée pourrait-elle faire ressortir des éléments de preuve non testés ou contestés sur « ce qui fonctionne, pour qui et où » ?
- Les conditions sont-elles favorables à l'évaluation pour qu'elle ait une influence ?  
En d'autres termes, comment les conclusions seront-elles utilisées ?  
Assurez-vous que vous êtes clairs à ce sujet de manière à vous concentrer sur l'utilisation.
- Le programme ou l'intervention est-il/elle prêt(e) pour une évaluation d'impact ? Par exemple, a-t-il/elle débuté depuis assez longtemps pour avoir des résultats ?
- Un budget suffisant est-il disponible pour les tailles d'échantillon nécessaires à la démonstration de l'impact au probable niveau d'effet (voir [Section 12 : L'échantillonnage](#)).



#### **Astuce :**

Après avoir traité les cinq questions ci-dessus, envisagez de réaliser une étude d'évaluabilité (voir [Section 2 : Décider de réaliser une évaluation](#)) avant de lancer une évaluation d'impact humanitaire nécessitant des ressources importantes.

#### **Que souhaitez-vous démontrer ou apprendre grâce à une évaluation d'impact ?**

Clarifier ce point vous aidera à définir le périmètre de l'évaluation d'impact. Par exemple, l'objectif est-il de savoir si l'intervention a eu un impact sur la population affectée de la manière souhaitée par les concepteurs et les bailleurs du programme ? Ou bien l'objectif porte-t-il sur l'impact plus large de l'intervention humanitaire, peut-être pour situer sa portée par rapport à ce que les gens ont fait de leur propre chef ?

#### **À l'intérieur du cycle du programme, quand est-il approprié de lancer une évaluation d'impact ?**

La réponse à cette question dépendra, en partie, de la réponse à la question précédente. Si vous vous intéressez à l'effet immédiat d'une intervention, il se peut que vous souhaitiez lancer votre évaluation d'impact quelques mois après le lancement de la mise en œuvre – par exemple, pour suivre et comprendre l'impact immédiat d'un programme d'abris d'urgence. Si vous vous intéressez aux conséquences et à l'impact à plus long terme de l'action humanitaire, il se peut que vous souhaitiez lancer votre évaluation d'impact plusieurs années après le commencement de la mise en œuvre – par exemple, pour explorer l'impact d'un programme conçu pour renforcer la résilience au niveau de la communauté et des ménages.

#### **Sera-t-elle plus judicieuse en tant qu'évaluation d'une seule organisation ou en tant qu'évaluation conjointe ?**

Les [évaluations conjointes d'impact](#) peuvent aider à garantir que l'impact général de l'action humanitaire est exploré plutôt que de tenter artificiellement d'isoler le travail



d'une seule organisation (Beck, 2009: 1). Elles permettent également la comparaison de différentes approches, ce qui peut aider à évaluer l'impact relatif de différentes interventions. Dans le secteur humanitaire, de nombreuses organisations qui réalisent des évaluations d'impact le font sous la forme d'exercices conjoints.

### **Quelles sont les questions d'évaluation appropriées ?**

Une fois clairement explicité ce que l'évaluation d'impact tente de découvrir ou d'apprendre, il est ensuite possible de structurer les questions d'évaluation pertinentes. Comme indiqué dans la [Section 6 : Choisir les questions d'évaluation](#), il est important d'identifier quelques questions de haut niveau conçues pour vous donner les réponses que vous cherchez. Rappelez-vous que vos questions d'évaluation doivent conduire le choix de la méthode et de la conception, et non l'inverse.

### **Quels sont les meilleures conceptions, méthodes et outils disponibles pour répondre à ces questions ?**

Pour répondre à cette question, soyez attentif aux caractéristiques du programme. Par exemple, le programme a-t-il une théorie du changement qui peut être testée ? Est-il possible de créer une hypothèse contrefactuelle ? Voir [Section 18.4](#).

### **Quels types de compétences évaluatives sont requis ?**

Une fois que le périmètre et les questions de l'évaluation sont définis, mais aussi que l'on s'est déjà fait une idée des méthodes les plus appropriées, la prochaine étape consiste à identifier les compétences évaluatives nécessaires – en particulier, si les compétences classiques sont adéquates ou s'il faut que les évaluateurs aient des compétences complémentaires en recherche qualitative et/ou quantitative. Proudlock *et al.* suggèrent que « la conception et la mise en œuvre d'évaluations d'impact nécessitent des compétences uniquement disponibles dans le cadre de partenariats à long terme entre universitaires, bailleurs, gouvernements, praticiens et bénéficiaires ciblés » (2009: 7).

### **Quelles sont les implications au niveau du budget et des ressources ?**

Parce que les évaluations d'impact minutieuses sont susceptibles de demander un travail de terrain plus long et intensif que d'autres types d'EAH, et peut-être des compétences supérieures, il est également probable qu'elles nécessitent des budgets plus importants qu'une EAH conventionnelle.



#### **Astuce :**

Les organisations doivent prioriser un petit nombre d'évaluations d'impact humanitaire de haute qualité et stratégique plutôt que de réaliser un grand nombre d'évaluations d'impact qui pourraient être sous-financées et qui, par conséquent, pourraient faire courir le risque de résultats de faible qualité.

## 18.4 Approches et conceptions

Les approches et méthodes qui suivent sont appropriées pour permettre à l'évaluation d'impact humanitaire de déduire des liens de causalité. Voir la [Section 11](#) pour plus de détails sur chacune de ces méthodes. Cette liste n'est pas exhaustive et d'autres méthodes possibles peuvent être envisagées. Des informations plus détaillées sont disponibles auprès d'initiatives menées par InterAction et le Département de la recherche de l'UNICEF - Innocenti<sup>3</sup>.

Approche/ conception	Potentiel	Contraintes	Astuces
<b>Essais contrôlés randomisés (RCT) ou conception de groupes de comparaison</b>	Lorsque cela est faisable, cela peut aider à établir une causalité en fournissant des données comparatives et une analyse crédible	Rarement faisable pour des raisons pratiques et éthiques dans des contextes humanitaires. Tests <i>ex-ante</i> rarement faisables, uniquement des tests <i>ex-post</i>	Envisagez des groupes de comparaison opportunistes, par exemple des camps qui ont reçu différents niveaux ou types d'assistance
<b>Les approches à base théorique</b>	Testez la théorie du changement sous-jacente d'une intervention de programme, par exemple en testant des hypothèses	La théorie du changement peut ne pas avoir été formulée par les concepteurs et les exécutants du programme	
<b>Les approches par cas</b>	Elles peuvent utiliser des méthodes qualitatives pour expliquer comment une intervention pourrait être responsable de changements particuliers, en s'appuyant sur un cas spécifique	Difficultés à généraliser à partir d'études de cas dans un contexte varié et évoluant rapidement	Les cas doivent être choisis de façon raisonnée
<b>Etude longitudinale</b>	Recherche sur une durée permettant de saisir les changements dans la vie de la population affectée	Susceptible de nécessiter des ressources importantes et pouvant ne pas être faisable dans des contextes où les groupes de population sont fortement mobiles en raison du conflit qui se prolonge, ou lorsque l'accès est intermittent	Il peut être possible d'utiliser des données issues d'une étude longitudinale existante – voir Khoo (2010) pour un exemple
<b>Approche dite de « contribution au changement » (Few <i>et al.</i>, 2014)</b>	Evalue les changements dans la durée dans la vie des populations affectées, dans quelle mesure leurs moyens de subsistance et leur bien-être ont été rétablis, ainsi que le rôle que les interventions semblent avoir joué dans ce processus de relèvement	Estime que les changements relatifs au bien-être et aux moyens de subsistance des personnes peuvent être le plus clairement identifiés au niveau des ménages. Principalement conçue pour être utilisée suite à des catastrophes naturelles à déclenchement rapide comme des crues soudaines, des tempêtes, des glissements de terrain, des séismes, des tsunamis et des éruptions volcaniques	

Une approche fondée sur les méthodes mixtes est particulièrement appropriée à l'évaluation d'impact humanitaire, surtout lorsque les méthodes mixtes se soutiennent mutuellement de façon synergique (Adato, 2011). Souvent, les méthodes mixtes sont les seules capables de gérer le grand nombre de facteurs potentiels de confusion dans le contexte habituel de l'action humanitaire. Le Feinstein Center a été le premier à utiliser une approche participative pour l'évaluation d'impact qui crée un partenariat d'apprentissage entre le bailleur et les partenaires de mise en œuvre (Catley *et al.*, 2009). L'exemple de bonnes pratiques à la page 365 démontre comment des méthodes qualitatives PRA peuvent être utilisées pour observer l'impact avec les communautés affectées de façon participative, même dans un environnement conflictuel difficile.



**Astuce :**

En choisissant la conception de votre évaluation d'impact, réfléchissez au niveau de certitude et de précision que l'évaluation doit atteindre dans ses réponses à des questions précises de manière à être jugée crédible. Cela aidera à renseigner l'approche d'échantillonnage, ainsi que tout compromis en termes de budget et de méthodes.



**Astuce :**

Dans les évaluations d'impact, il est crucial de réaliser une consultation étendue et approfondie de la population affectée pour saisir son point de vue. L'évaluation d'impact humanitaire dépend presque toujours, dans une certaine mesure, d'un jugement de valeur, ce qui signifie que le jugement de la population affectée doit être prépondérant pour déterminer la causalité.



### Exemple de bonnes pratiques : La série d'évaluations d'impact conjointes du PAM et du HCR sur la contribution de l'assistance alimentaire aux solutions durables pour des réfugiés en situation prolongée

Le PAM et le HCR ont conjointement commandité et conduit une série d'évaluations d'impact utilisant des méthodes mixtes au Bangladesh, au Tchad, en Ethiopie et au Rwanda, pour évaluer la contribution de l'assistance alimentaire aux solutions durables pour des réfugiés en situation prolongée (PAM et HCR, 2012). Globalement, cette série d'évaluations d'impact a utilisé une approche à base théorique pour déduire des liens de causalité. Les quatre évaluations ont utilisé le même cadre théorique et la même approche, des détails étant adaptés pour chaque contexte. Parmi les caractéristiques clés de cette série d'évaluations d'impact, on compte notamment son utilisation d' :

- Un modèle logique (théorie du changement, TdC) qui a d'abord été développé par le Bureau de l'évaluation du PAM, puis discuté et validé par les équipes d'évaluation dans les quatre pays visités ; les équipes ont également évalué l'adéquation de la TdC avec les cadres logiques au niveau pays et ce, aux différentes phases de l'évaluation.
- Une approche utilisant des méthodes mixtes et adaptée à une situation où il n'était pas possible d'utiliser une hypothèse contrefactuelle conventionnelle comme base pour déduire des liens de causalité. L'utilisation de méthodes mixtes pour rassembler et analyser les données des quatre études de cas pays comprenait une triangulation des données issues des revues documentaires ; des entretiens avec des parties prenantes interviews du PAM et du HCR ; des revues de données secondaires ; des enquêtes quantitatives ; des marches d'observation ; et des entretiens qualitatifs, notamment avec des *focus groups* de bénéficiaires et des membres de communautés locales accueillant des réfugiés (PAM et HCR, 2012: 2).
- Une analyse évaluative qui s'appuie sur les résultats émergeant des études de cas pays pour déterminer : a) quels facteurs internes et externes pourraient expliquer les résultats en termes de causalité ; et b) quels facteurs ont influencé les résultats et les changements (intentionnels et non intentionnels) observés dans les différents pays et pourquoi (voir PAM et HCR, 2012: 10-13).

Source : Alexander et Bonino (2015: 11)



### **Exemple de bonnes pratiques : Evaluations d'impact participatives**

Pour un bon exemple d'évaluation d'impact participative, voir la distribution de kits de moyens de subsistance d'urgence par la FAO au Soudan du Sud – avril-mai 2015.

Dans le cadre de l'évaluation du programme de la FAO au Soudan du Sud (focalisée sur la réponse au niveau d'urgence 3), le Bureau d'évaluation de la FAO a réalisé une évaluation d'impact participative dans certaines zones les plus gravement affectées par le conflit en 2014, lorsque la FAO est intervenue en distribuant des kits de moyens de subsistance qui contenait des semences de culture vivrière, des outils et du matériel de pêche.

L'évaluation a cherché à collecter, auprès des communautés ciblées, des informations de première main sur les impacts positifs, négatifs, intentionnels ou non de la distribution, cela dans le but d'améliorer les prochaines interventions d'urgence et de recevoir des commentaires de la part des personnes affectées. Des méthodes de recherche participatives ont été utilisées, en particulier des outils qualitatifs PRA. Des cartes de ressources des villages et les calendriers ont été élaborés par des *focus groups* distincts d'hommes et de femmes de façon à comprendre les changements au niveau des actifs et de la vulnérabilité des communautés. Des matrices portant sur les revenus, les dépenses et les stratégies de survie, ainsi que des plannings journaliers d'activités ont été utilisés pour évaluer l'impact de la distribution au niveau des ménages, mais aussi les questions de nutrition. Trois zones géographiques différentes ont été visitées afin d'inclure des lieux sous contrôle du gouvernement mais aussi de l'opposition.

Une équipe constituée de personnel national avec des origines ethniques variées a été choisie, ce qui a permis à ces évaluateurs d'avoir accès aux différentes zones. Ils ont ensuite été formés aux méthodes PRA et on leur a demandé de tester ces méthodologies dans une zone « sûre ». Enfin, ils ont développé un manuel de formation aux méthodes PRA à destination des enquêteurs locaux. Les résultats de cette évaluation ont été intégrés dans le programme d'urgence 2016 de la FAO au Soudan du Sud.

Source : (FAO, mimeo).

Voir également l'[Exemple des bonnes pratiques à la page 290](#) dans la [Section 15 : L'accès restreint](#), pour une brève description d'une étude d'impact réalisée en Somalie dans le cadre de l'évaluation 2012/13 de la coopération de la FAO en Somalie entre 2007 et 2012.

## Notes de fin de texte

1. Voir Alexander et Bonino (2015) ainsi que Morra Imas et Rist (2009).
2. Tiré de Bonino (2015), qui est résumé dans le guide protection, d'après Rogers (2012) et Chigas, Church et Corlazzoli (2014).
3. InterAction a développé un Guide d'évaluation d'impact et une série de webinars : [www.interaction.org/impact-evaluation-notes](http://www.interaction.org/impact-evaluation-notes).

Le Département de la recherche de l'UNICEF – Innocenti a collaboré avec l'Université RMIT, BetterEvaluation et l'International Initiative for Impact Evaluation (3ie) pour développer une série de notes méthodologiques et de vidéos sur l'évaluation d'impact : [www.unicef-irc.org/KM/IE/impact.php](http://www.unicef-irc.org/KM/IE/impact.php).

# Glossaire

## **Action humanitaire / 24**

Les objectifs de l'action humanitaire sont de sauver des vies, d'atténuer les souffrances et de préserver la dignité humaine pendant et après des crises provoquées par l'homme ou des catastrophes naturelles, ainsi que de prévenir de telles situations et d'améliorer la préparation à leur survenue.

## **Analyse de contenu / 180**

Analyse de l'information textuelle d'une manière standardisée permettant aux évaluateurs de procéder à des déductions à partir de l'information.

## **Analyse des mots-clés / 181**

Forme d'analyse de contenu qui examine la fréquence de l'occurrence des catégories pour mettre en évidence les tendances au fil du temps dans une seule série de documents ou pour comparer deux séries de documents.

## **Appariement par score de propension / 198**

Technique d'appariement statistique qui cherche à associer le groupe de comparaison avec le groupe de contrôle en choisissant un qui ait la même probabilité de recevoir une aide selon ses caractéristiques.

## **Apprentissage / 27**

Processus par lequel l'expérience et la réflexion mènent à des changements de comportement ou à l'acquisition de nouvelles capacités.

## **Attribution / 29**

L'imputation d'un lien de causalité entre des changements observés (ou que l'on s'attend à observer) et une intervention spécifique.

## **Auto-évaluation / 85**

Évaluation réalisée par ceux qui conçoivent et mettent en œuvre une intervention, synonyme d'évaluation interne.

## **Codage / 304**

Attribue des catégories à des éléments de preuve précis.

## **Comparaison avant/après / 204**

Une comparaison avant/après compare la situation d'un groupe avant et après l'intervention.

## **Comparaison de discontinuité du traitement / 206**

compare le groupe situé juste sous le point d'exclusion de l'assistance avec le groupe situé juste au-dessus.

## **Conception de discontinuité de la régression / 205**

Compare les lignes de régression pour la variable d'intérêt avec la note sur laquelle s'est basée l'intervention.

## **Conceptions expérimentales / 208**

Sont celles où les unités d'analyse sont déterminées de façon aléatoire au groupe recevant l'assistance ou au groupe de contrôle. Chaque élément (par exemple, une personne, une famille ou une communauté) a une chance égale d'être désigné groupe assisté ou groupe de contrôle.

## **Conceptions en série chronologique interrompue / 203**

Fournit une estimation de l'impact d'une intervention en examinant une série chronologique de données avant et après une intervention.

**Conception participative / 202**

Implique toutes les parties prenantes à toutes les phases de l'évaluation, de la planification initiale à la mise en œuvre des recommandations.

**Conceptions non-expérimentales / 195**

Conceptions qui n'ont recours à aucune comparaison, que ce soit entre populations assistées et populations non-assistées, ou au fil du temps au sein des populations qui ont reçu une assistance.

**Conceptions quasi-expérimentales / 197**

Conceptions qui utilisent une comparaison lorsque le groupe de comparaison n'est pas choisi de manière aléatoire.

**Conclusion / 319**

Affirmation inductive fondée sur une ou plusieurs constatations.

**Constatation (traduction de « finding ») / 319**

Affirmation factuelle fondée sur un élément de preuve.

**Contribution / 29**

Analyser la contribution dans le cadre de l'évaluation revient à montrer de façon crédible qu'une intervention a joué un certain rôle dans l'obtention de résultats. L'analyse de la contribution est une sorte d'analyse évaluative qui reconnaît que plusieurs causes peuvent contribuer à un résultat, même si individuellement elles peuvent ne pas être nécessaires ou suffisantes pour avoir un impact.

**Crowd-sourcing (approvisionnement par la foule) / 286**

A recours à un grand nombre de volontaires soit pour collecter des données, soit pour analyser des données d'imagerie, généralement à l'aide de moyens indirects. Ce type de données est appelée données crowd-sourcées.

**Diffusion / 339**

Processus visant à communiquer des informations à des publics définis dans le but d'augmenter les connaissances mais aussi de modifier les politiques et les pratiques.

**Données primaires / 293**

Données collectées pour les besoins de l'évaluation.

**Données secondaires / 293**

Données collectées pour d'autres besoins mais utilisées par l'évaluation.

**Ecart dans les différences / 204**

Cette conception estime l'effet de l'assistance en comparant dans le temps l'évolution de la moyenne du résultat d'intérêt entre le groupe assisté et un groupe de comparaison.

**Ecart standard / 228**

Mesure de la variabilité d'un paramètre.

**Echantillonnage / 214**

Sélection d'un sous-ensemble de population à prendre en compte dans une étude (plutôt que la population dans son ensemble).

**Echantillonnage aléatoire / 205**

Etablit un échantillon d'une population où chaque membre de la population a une chance égale d'être sélectionné.



**Echantillonnage de disponibilité / 217**

Procédure d'échantillonnage où la sélection est fondée sur la disponibilité des échantillons, la convenance du chercheur ou l'auto-sélection.

**Echantillonnage par grappes / 225**

Echantillonnage où un certain nombre de sites sont échantillonnés, chacun comportant un nombre spécifique de cas.

**Echantillonnage non aléatoire / 214**

Choisit l'échantillon d'après certaines propriétés de l'échantillon.

**Echantillonnage par quotas 219**

Divise la population en sous-catégories qui s'excluent mutuellement, et collecte ensuite les informations selon une taille d'échantillon précédemment établie ou une proportion pour chaque catégorie.

**Echantillonnage pseudo-aléatoire / 224**

Echantillonnage retenu lorsqu'il n'existe aucun cadre d'échantillonnage. De manière générale, le premier choix est sélectionné de façon aléatoire à partir d'un point de départ sélectionné de façon raisonnée, et les choix suivants sont sélectionnés en utilisant la même règle.

**Echantillonnage raisonné 218**

Sélectionne l'échantillon sur une base intentionnelle de façon à ce que les éléments échantillonnés puissent fournir le plus d'informations possibles à l'étude.

**Echantillonnage stratifié / 226**

Approche d'échantillonnage où la population est divisée en segments s'excluant mutuellement avant qu'un simple échantillon aléatoire ne soit sélectionné au sein de chacun d'eux.

**Effet de conception / 225**

(également appelé « effet de plan ou effet du plan de sondage ») Facteur par lequel vous devez modifier votre taille d'échantillon lorsque vous vous écarterez d'un simple échantillonnage aléatoire.

**Éléments de preuve / 293**

Ensemble des faits ou informations disponibles capables d'étayer une proposition ou une idée particulière.

**Essai contrôlé randomisé / 207**

Compare deux groupes sélectionnés de façon aléatoire, l'un des deux recevant l'assistance mais pas l'autre (le groupe de contrôle).

**Etude de cas / 200**

Description et une analyse intensive d'un ou plusieurs cas (qui peuvent être des individus comme des Etats) permettant de tirer des conclusions générales sur l'intervention.

**Etude de processus / 201**

Compare la manière dont fonctionnent les processus avec le fonctionnement prévu.

**Etude de résultats / 202**

Compare les résultats avec les résultats prévus.

**Etudes longitudinale / 206**

Les études longitudinales effectuent des mesures répétées de la même population sur plusieurs années.

**Évaluand (Objet de l'évaluation) / 210**

Le sujet d'une évaluation, en général un programme ou un système plutôt qu'une personne.

**Évaluation à mi-parcours / 83**

Évaluation réalisée à la moitié d'une intervention.

**Evaluation au sein d'une même organisation / 84**

Evaluation menée par l'organisation qui a mis en place l'intervention.

**Evaluation conjointe / 84**

Evaluation menée par deux organisations ou plus, évaluant le travail de deux organisations ou plus.

**Evaluation d'impact / 81**

Evaluation qui se concentre sur les effets plus larges d'un programme humanitaire, notamment l'impact intentionnel et non intentionnel, l'impact positif et négatif, et enfin l'impact macro (sectoriel) et micro (ménage, individu).

**Evaluation d'un portfolio humanitaire / 81**

Evaluation du portfolio humanitaire complet d'une organisation.

**Evaluation de Cluster / 80**

Evaluation de multiples projets au sein d'un programme plus vaste, OU évaluation en lien avec le système de coordination par Cluster des Nations unies.

**Evaluation de l'action humanitaire (EAH) / 27**

Examen systématique et objectif de l'action humanitaire afin de déterminer la valeur ou la signification d'une activité, d'une politique ou d'un programme, cela dans le but de tirer des leçons permettant d'améliorer la politique et la pratique, ainsi que d'augmenter la redevabilité.

**Evaluation de l'ensemble du système / 84**

Evaluation de la réponse du système humanitaire international à une crise humanitaire, ouverte à tous les acteurs du système.

**Evaluations externes ou indépendantes / 85**

Evaluation réalisées par des évaluateurs n'appartenant pas à l'équipe qui a mis en œuvre le programme.

**Evaluation de partenaire / 80**

Evaluation d'un ensemble d'interventions mises en œuvre par un seul partenaire.

**Evaluation de processus / 82**

Evaluation qui se concentre sur les processus par lesquels les moyens sont convertis en produits. Elle peut aussi examiner l'intervention dans son ensemble.

**Evaluation de programme / 79**

Evaluation d'un ensemble d'interventions présentant un objectif humanitaire partagé.

**Evaluation de projet / 79**

Evaluation d'une unique intervention humanitaire (avec des objectifs, des ressources et un calendrier de mise en œuvre spécifiques) qui se déroule souvent dans le cadre d'un programme plus vaste.

**Evaluation des politiques / 86**

Evaluation qui examine les compréhensions, croyances et hypothèses qui rendent les projets individuels possibles et souhaitables. Elle peut évaluer à la fois l'efficacité d'une politique et la manière dont la politique a été mise en œuvre.

**Evaluation en temps réel (ETR) / 82**

Evaluation d'une opération humanitaire en cours à mesure qu'elle se déroule.

**Evaluation ex-ante / 83**

Evaluation conduite avant qu'une intervention ne commence.

**Evaluation ex-post / 83**

Evaluation réalisée à la fin d'une intervention.

**Evaluation institutionnelle / 86**

Evaluation des dynamiques internes des organisations chargées de la mise en œuvre de l'aide, de leurs instruments politiques, de leurs mécanismes d'acheminement des services et de leurs pratiques de gestion ainsi que des liens entre ces éléments.

**Evaluation itérative / 83**

Série d'évaluations conçue pour se dérouler tout au long d'une intervention.

**Evaluation normative / 82**

Evaluation qui compare ce qui est mis en œuvre avec ce qui était planifié ou avec des normes spécifiques.

**Evaluation participative / 85**

Evaluation durant laquelle les parties prenantes, y compris la population affectée, coopèrent pour concevoir, mener et interpréter une évaluation.

**Evaluation participative rapide (EPR) / 276**

Les techniques de l'évaluation participative rapide (EPR) (également appelée « évaluation rurale participative ») est un groupe de méthodes permettant aux populations locales de renforcer, d'analyser et de partager leurs connaissances et leur apprentissage de façon à ce que des chercheurs venus de l'extérieur puissent les comprendre facilement.

**Evaluation sectorielle / 80**

« Evaluation d'un groupe d'interventions dans un secteur, lesquelles contribuent toutes à atteindre un objectif humanitaire spécifique. L'évaluation peut couvrir une partie d'un pays, un pays entier, ou plusieurs pays » (UNICEF, 2013).

**Evaluation technologique / 86**

Evaluation d'une technique ou d'une technologie spécifique.

**Evaluation thématique / 81**

Evaluation portant sur une sélection d'interventions qui répondent toutes à une priorité humanitaire spécifique commune à plusieurs pays, régions, et éventuellement organisations et secteurs.

**Expérimentation naturelle / 208**

Compèrent un groupe assisté et un groupe similaire qui, par hasard, n'a pas reçu d'assistance.

**Grille d'évaluation / 178**

Outil de notation utilisé pour évaluer un document à partir d'une série de critères de façon harmonisée.

**Groupe de comparaison / 205**

Les conceptions de groupe de comparaison compèrent le groupe assisté avec un groupe de comparaison sélectionné.

**Groupe de gestion / 150**

Groupe qui gère l'évaluation au jour le jour et dont les tâches comprennent la rédaction des Termes de Référence, le recrutement et la gestion de l'équipe d'évaluation, la gestion de la lecture et la finalisation du rapport d'évaluation.

**Groupe de pilotage / 149**

Groupe mis en place pour piloter une évaluation à travers ses étapes clés comme la rédaction des Termes de Référence, du rapport initial et du rapport final.

**Groupe de référence / 150**

Groupe constitué des principales parties prenantes connaissant bien l'environnement local et pouvant apporter des conseils sur des questions pratiques en lien avec l'évaluation et sur la faisabilité des recommandations qui en découlent.

**Groupe de revue par des pairs / 151**

Groupe qui conseille sur les questions de qualité, généralement composé d'évaluateurs et d'autres spécialistes choisis pour leurs connaissances de l'évaluation, de la région ou du type d'intervention évalué.

**Impact / 29**

S'intéresse aux effets plus larges du programme – sociaux, économiques, techniques et environnementaux – sur les individus, les groupes de genre et d'âge, les communautés et les institutions. Les impacts peuvent être intentionnels et non intentionnels, positifs et négatifs, macro (sectoriels) et micro (ménage, individu), à court ou long terme.

**Inférence causale / 302**

Etablissement d'une relation entre une cause et un effet.

**Intervalle de confiance / 227**

Fourchette à l'intérieur de laquelle nous nous attendons à trouver la valeur de la population dans son ensemble.

**Méta-évaluation / 87**

Méta-évaluation est une évaluation globale conçue pour agréger des résultats issus d'une série d'évaluations, généralement en fonction d'un ensemble de critères. Elle peut également être utilisée pour désigner l'« évaluation d'une évaluation », c'est-à-dire pour juger la qualité d'une série d'évaluations.

**Modèle logique / 92**

Tableau ou diagramme présentant la théorie du programme (manière dont les moyens sont supposés contribuer à l'objectif général d'une intervention).

**Moyens / 28**

Les ressources financières, humaines et matérielles utilisées dans l'action humanitaire.

**Niveau de confiance / 227**

Probabilité que la valeur de la population dans son ensemble ne tombe dans l'intervalle de confiance.

**Outil d'enquête / 233**

Questionnaire utilisé par l'intervieweur durant les entretiens d'enquêtes formelles.

**Phase de démarrage / 135**

La phase de démarrage de l'évaluation va de la sélection de l'équipe d'évaluation à l'acceptation du rapport de démarrage.

**Produits / 28**

Constructions, biens et services qui résultent d'une intervention.

**Protection / 25**

« La protection comprend l'ensemble des activités destinées à faire pleinement respecter les droits des personnes, conformément à la lettre et à l'esprit des branches pertinentes du droit » (IASC, 2011).

**Puissance statistique / 228**

Probabilité qu'un résultat négatif ne soit pas un résultat négatif faux (1 moins le risque d'un résultat négatif faux).

**Raisonnement évaluatif / 297**

Processus analytique par lequel les évaluateurs répondent aux questions évaluatives.

**Recommandation / 312**

Façon de procéder que les évaluateurs suggèrent comme manière de traiter une ou plusieurs conclusions.

**Redevabilité / 27**

Moyen par lequel le pouvoir est utilisé de façon responsable. Elle consiste en un processus qui prend en compte les points de vue des différentes parties prenantes, à être tenu responsable devant celles-ci, et en premier lieu devant les populations affectées par les autorités ou le pouvoir en place.

**Résultats (traduction de « outcomes ») / 283**

Changements ou transformations intentionnel(le)s ou non intentionnel(le)s de la situation causé(e)s directement ou indirectement par une intervention. Ceux-ci peuvent être souhaités (positifs) ou indésirables (négatifs). Ils englobent les changements de comportement (actions, relations, politiques, pratiques) des individus, groupes, communautés, organisations, institutions ou autres acteurs sociaux.

**Revue après action / 265**

Echange structuré à propos d'une intervention humanitaire qui permet à une équipe de réfléchir à ce qui s'est passé, à pourquoi cela s'est passé et à la manière de conserver les points forts et d'améliorer les faiblesses.

**Revue thématique / 350**

Extrait les constatations et les leçons apprises d'une série d'évaluations de l'action humanitaire dans un sous-secteur, par exemple la protection, l'habitat/abri. Ce type d'étude peut être commandité pour répondre / traiter une série précise de questions fondamentales.

**Saturation des données / 217**

La saturation des données se produit lorsque de nouveaux cas n'apportent plus de nouvelles connaissances.

**Suivi / 30**

Processus continu recourant à la collecte systématique de données à partir d'indicateurs spécifiés pour fournir des indications sur l'étendue des progrès, l'atteinte des objectifs et le niveau de consommation des fonds alloués aux gestionnaires et aux principales parties prenantes d'une intervention humanitaire en cours. (D'après OCDE-CAD, 2002.)

**Statistiques descriptives / 312**

Sont utilisées pour résumer des aspects essentiels d'une population.

**Statistiques inférentielles / 312**

Sont utilisées soit pour tirer des déductions sur une population à partir d'un échantillon, soit pour tirer des déductions sur des hypothèses.

**Synthèse des évaluations / 349**

Suit une procédure systématique pour organiser et résumer les résultats d'une série d'évaluations (potentiellement assez disparates). Cela peut s'avérer une opportunité pour poser des questions précises sur la série d'évaluations, et/ou pour identifier des tendances courantes dans les résultats.

**Tableau de gestion des recommandations / 347**

Registre de la réponse associée à chaque recommandation de l'évaluation ainsi que des étapes que les dirigeants prévoient de suivre pour y répondre, comprenant une date butoir et la partie responsable de chaque étape.

**Tableau évaluatif / 297**

Tableau qui décrit à quoi devraient ressembler les éléments de preuve à différents niveaux de performance, pour certains critères d'intérêt ou pour l'intervention dans son ensemble.

**Taille de l'effet / 228**

Différence proportionnée entre la variable étudiée du groupe traité et celle du groupe de contrôle.

**Termes de Référence / 118**

Présentent « une vue d'ensemble des exigences et des attentes de l'évaluation. Ils fournissent un énoncé explicite des objectifs de l'évaluation, des rôles et des responsabilités des évaluateurs et du client de l'évaluation ainsi que des ressources disponibles pour l'évaluation » (Roberts *et al.*, 2011 : 2).

**Théorie du changement / 97**

Description du mécanisme central par lequel le changement se produit pour les individus, les groupes et les communautés.

**Utilisation / 62**

Sandison, en s'appuyant les travaux de Patton (1997) et Carlsson *et al.* (1999), propose la définition suivante de l'utilisation : « une évaluation a été utilisée si des utilisateurs ayant l'intention et le potentiel pour agir ont pris en compte de façon sérieuse et active ses conclusions, identifiant des usages pertinents par rapport à leurs propres intérêts et besoins » (2006 : 100-101).

## Bibliographie

Les publications suivantes sont disponibles via le Portail de l'évaluation humanitaire et de l'apprentissage (HELP) : [www.alnap.org/resources/EHA-2016](http://www.alnap.org/resources/EHA-2016)

**Abate, G. T., de Brauw, A., Minot, N. and Bernard, T. (2014)** *The impact of the use of new technologies on farmers wheat yield in Ethiopia: Evidence from a randomized controlled trial*. Washington, DC: REAP. ([www.alnap.org/resource/22846.aspx](http://www.alnap.org/resource/22846.aspx)).

**Abdullah, M. Y. H., Bakar, N. R., Sulehan, J., Awang, A. H. and Liu, O. P. (2012)** 'Participatory rural appraisal (PRA): An analysis of experience in Darmareja village, Sukabumi District, West Java, Indonesia'. *Akademika*, 82(1): 15–19. ([www.alnap.org/resource/22847.aspx](http://www.alnap.org/resource/22847.aspx)).

**ACAPS. (2011)** *Technical brief: Direct observation and key informant interview techniques for primary data collection during rapid assessments*. Geneva: ACAPS. ([www.alnap.org/resource/22848.aspx](http://www.alnap.org/resource/22848.aspx)).

**ACAPS. (2013)** *How sure are you? Judging quality and usability of data collected during rapid needs assessments*. Geneva: ACAPS. ([www.alnap.org/resource/11437.aspx](http://www.alnap.org/resource/11437.aspx)).

**ACF. (2014)** *ACF Learning Review 2014*. Paris: ACF. ([www.alnap.org/resource/20231.aspx](http://www.alnap.org/resource/20231.aspx)).

**Adato, M. (2011)** *Combining quantitative and qualitative methods for program monitoring and evaluation: Why are mixed-method designs best?* Washington, DC: The World Bank. ([www.alnap.org/resource/8069.aspx](http://www.alnap.org/resource/8069.aspx)).

**Adinolfi, C., Bassiouni, D., Lauritzsen, H. and Williams, H. (2005)** *Humanitarian Response Review*. New York: UN OCHA. ([www.alnap.org/resource/3284.aspx](http://www.alnap.org/resource/3284.aspx)).

**Ager, A., Ager, W., Stavrou, V. and Boothby, N. (2011)** *Inter-agency guide to the evaluation of psychosocial programming in humanitarian crises*. New York: UNICEF. ([www.alnap.org/resource/10214.aspx](http://www.alnap.org/resource/10214.aspx)).

**Ager, A., Stark, S. and Potts, A. (2010)** *Participative ranking methodology: A brief guide: Version 1.1*. New York: Columbia University. ([www.alnap.org/resource/8070.aspx](http://www.alnap.org/resource/8070.aspx)).

**Ager, A., Stark, L., Sparling, T. and Ager, W. (2011)** *Rapid appraisal in humanitarian emergencies using participatory ranking methodology (PRM)*. New York: Columbia University. ([www.alnap.org/resource/22849.aspx](http://www.alnap.org/resource/22849.aspx)).

**Ahmed, L., Malik, Z. and Nawab, B. (2012)** *Humanitarian support to conflict and flood-affected populations in Pakistan*. Paris: ACF. ([www.alnap.org/resource/21080.aspx](http://www.alnap.org/resource/21080.aspx)).

**Alexander, J. (2014)** *Improving accuracy in humanitarian evaluations*. London: ALNAP. ([www.alnap.org/resource/12636.aspx](http://www.alnap.org/resource/12636.aspx)).

**Alexander, J. and Bonino, F. (2014)** *Ensuring quality of evidence generated through participatory evaluation in humanitarian context*. London: ALNAP. ([www.alnap.org/resource/19163.aspx](http://www.alnap.org/resource/19163.aspx)).

**Alexander, J. and Bonino, F. (2015)** *Addressing causation in humanitarian evaluation: A discussion on designs, approaches and examples*. London: ALNAP. ([www.alnap.org/resource/19453.aspx](http://www.alnap.org/resource/19453.aspx)).

**Alexander, J. and Cosgrave, J. (2014b)** *Representative sampling in humanitarian evaluation*. London: ALNAP. ([www.alnap.org/resource/10389.aspx](http://www.alnap.org/resource/10389.aspx)).

**Alexander, J., Darcy, J. and Kiani, M. (2013)** *The 2013 humanitarian accountability report*. Geneva: HAP. ([www.alnap.org/resource/8758.aspx](http://www.alnap.org/resource/8758.aspx)).

**ALNAP. (2002)** *ALNAP annual review 2002: Humanitarian action: improving performance through improved learning*. London: ALNAP. ([www.alnap.org/resource/5189.aspx](http://www.alnap.org/resource/5189.aspx)).

**ALNAP. (2005)** *Assessing the quality of humanitarian evaluations. The ALNAP quality proforma 2005*. London: ALNAP. ([www.alnap.org/resource/5320.aspx](http://www.alnap.org/resource/5320.aspx)).

**ALNAP. (2007a)** *An introduction to evaluation of humanitarian action (EHA) course manual*. London: ALNAP. ([www.alnap.org/resource/22850.aspx](http://www.alnap.org/resource/22850.aspx)).

**ALNAP. (2007b)** *Slow-onset disasters: drought and food and livelihoods insecurity—learning from previous relief and recovery responses*. London: ALNAP. ([www.alnap.org/resource/5243.aspx](http://www.alnap.org/resource/5243.aspx)).

**ALNAP. (2009)** *Participation by crisis-affected populations in humanitarian action. A handbook for practitioners*. London: ALNAP. ([www.alnap.org/resource/5271.aspx](http://www.alnap.org/resource/5271.aspx)).



- ALNAP. (2015)** *The state of the humanitarian system*. London: ALNAP. ([www.alnap.org/resource/21036.aspx](http://www.alnap.org/resource/21036.aspx)).
- ALNAP and DEC. (2013)** *Workshop Summary. Syria Crisis: Evaluators' learning exchange*. London: ALNAP. ([www.alnap.org/resource/9300.aspx](http://www.alnap.org/resource/9300.aspx)).
- ALNAP and Groupe URD. (2009)** *Participation handbook for humanitarian field workers: Involving crisis-affected people in humanitarian response*. London and Plaisians: ALNAP and Groupe URD. ([www.alnap.org/resource/8531.aspx](http://www.alnap.org/resource/8531.aspx)).
- ALNAP, OECD-DAC, and UNEG. (2010)** *Supporting Evaluation in Haiti Mission Report*. London: ALNAP. ([www.alnap.org/resource/22851.aspx](http://www.alnap.org/resource/22851.aspx)).
- Amazon Fund. (2010)** *Logical framework*. Amazon Fund. ([www.alnap.org/resource/22852.aspx](http://www.alnap.org/resource/22852.aspx)).
- AEA. (2003)** *American Evaluation Association response to US Department of Education*. Washington, DC: AEA. ([www.alnap.org/resource/22855.aspx](http://www.alnap.org/resource/22855.aspx)).
- AEA. (2004)** *Guiding principles for evaluators*. Washington, DC: AEA. ([www.alnap.org/resource/22853.aspx](http://www.alnap.org/resource/22853.aspx)).
- Anderson, A.A. (2005)** *The community builder's approach to theory of change: a practical guide to theory development*. New York: The Aspen Institute. ([www.alnap.org/resource/22854.aspx](http://www.alnap.org/resource/22854.aspx)).
- Ashdown, P. (2011)** *Humanitarian emergency response review*. London: DFID. ([www.alnap.org/resource/6355](http://www.alnap.org/resource/6355)).
- Audsley, B., Halme, R. and Balzer, N. (2010)** 'Comparing cash and food transfers: a cost-benefit analysis from rural Malawi', in Were Omamo, S., Gentilini, U. and Sandström, S. (eds) *Revolution: From food aid to food assistance: Innovations in overcoming hunger*. Rome: WFP. ([www.alnap.org/resource/22856.aspx](http://www.alnap.org/resource/22856.aspx)).
- AusAID. (2013)** *Shelter sector response evaluation typhoon Pablo December 2012 in Mindanao, Philippines - shelter cluster report*. Canberra: AusAID. ([www.alnap.org/resource/10074.aspx](http://www.alnap.org/resource/10074.aspx)).
- Baez, J.E. and Santos, I.V. (2007)** *Children's vulnerability to weather shocks: A natural disaster as a natural experiment*. New York: Social Science Research Network. ([www.alnap.org/resource/8078.aspx](http://www.alnap.org/resource/8078.aspx)).

- Bailey, S. (2013)** *Evaluation of Concern Worldwide's emergency response in Masisi, North Kivu, DRC (2012-2013)*. London: Concern. ([www.alnap.org/resource/22857.aspx](http://www.alnap.org/resource/22857.aspx)).
- Baker, J., Chantal, S., Hidalgo, S., Kayungura, G., Posada, S., Tasikasereka, M. and Vinas, M. (2013)** *External evaluation of the rapid response to population movements (RRMP) program in the Democratic Republic of Congo*. Madrid: DARA. ([www.alnap.org/resource/12494.aspx](http://www.alnap.org/resource/12494.aspx)).
- Bamberger, M. (2012)** *Introduction to mixed methods in impact evaluation*. Washington, DC: InterAction. ([www.alnap.org/resource/8079.aspx](http://www.alnap.org/resource/8079.aspx)).
- Bamberger, M., Carden, F. and Rugh, J. (2009)** *Alternatives to the conventional counterfactual: summary of session 713 think tank: American Evaluation Association, Orlando 2009*. Washington, DC: AEA. ([www.alnap.org/resource/8222.aspx](http://www.alnap.org/resource/8222.aspx)).
- Bamberger, M., Rao, V. and Woolcock, M. (2010)** *Using mixed methods in monitoring and evaluation: experiences from international development*. Washington, DC: The World Bank. ([www.alnap.org/resource/22281.aspx](http://www.alnap.org/resource/22281.aspx)).
- Bamberger, M., Rugh, J. and Mabry, L. (2011)** *RealWorld evaluation. Working under budget, time, data, and political constraints: A condensed summary overview. 2nd ed.* London: SAGE. ([www.alnap.org/resource/8076.aspx](http://www.alnap.org/resource/8076.aspx)).
- Bamberger, M. and Segone, M. (2011)** *How to design and manage equity-focused evaluations*. New York: UNICEF. ([www.alnap.org/resource/8080.aspx](http://www.alnap.org/resource/8080.aspx)).
- Barahona, C. (2010)** *Randomised control trials for the impact evaluation of development initiatives: a statistician's point of view*. Rome: ILAC. ([www.alnap.org/resource/22858.aspx](http://www.alnap.org/resource/22858.aspx)).
- Barakat, S. (2006)** *Mid-term evaluation report of the national solidarity programme (NSP), Afghanistan*. York: University of York. ([www.alnap.org/resource/22859.aspx](http://www.alnap.org/resource/22859.aspx)).
- Barakat, S., Hardman, F., Connolly, D., Sundaram, V. and Zyck, S. A. (2010)** *Programme review and evaluability study (PRES): UNICEF's education in emergencies and post-crisis transition (EEPCT) programme*. New York: UNICEF. ([www.alnap.org/resource/8081.aspx](http://www.alnap.org/resource/8081.aspx)).

- Barron, I. G., Abdallah, G. and Smith, P. (2012)** 'Randomized control trial of a CBT trauma recovery program in Palestinian schools'. *Journal of Loss and Trauma*, 2013(18): 306-321. ([www.alnap.org/resource/22860.aspx](http://www.alnap.org/resource/22860.aspx)).
- Baugh, J. B., Hallcom, A. S. and Harris, M. E. (2010)** 'Computer assisted qualitative data analysis software: A practical perspective for applied research'. *Revista del Instituto Internacional de Costos*. ([www.alnap.org/resource/22861.aspx](http://www.alnap.org/resource/22861.aspx)).
- Beall, J. and Schütte, S. (2006)** *Urban livelihoods in Afghanistan*. Kabul: AREU. ([www.alnap.org/resource/22862.aspx](http://www.alnap.org/resource/22862.aspx)).
- Beck, T. (2006)** *Evaluating humanitarian action using the OECD-DAC criteria*. ALNAP Guide. London: ALNAP. ([www.alnap.org/resource/5253.aspx](http://www.alnap.org/resource/5253.aspx)).
- Beck, T. (2009)** *Joint humanitarian impact evaluation: options paper*. London: ALNAP. ([www.alnap.org/resource/5741.aspx](http://www.alnap.org/resource/5741.aspx)).
- Beck, T. (2011)** *Joint humanitarian impact evaluation: report on consultations: Report for the inter-agency working group on joint humanitarian impact evaluation*. London: ALNAP. ([www.alnap.org/resource/6046.aspx](http://www.alnap.org/resource/6046.aspx)).
- Beck, T. and Buchanan-Smith, M. (2008)** 'Joint evaluations coming of age? The quality and future scope of joint evaluations', in *ALNAP 7th Review of Humanitarian Action*. ALNAP Review. London: ALNAP. ([www.alnap.org/resource/5232.aspx](http://www.alnap.org/resource/5232.aspx)).
- Benatar, S. R. (2002)** 'Reflections and recommendations on research ethics in developing countries'. *Social Science & Medicine*, 54(7). ([www.alnap.org/resource/8084.aspx](http://www.alnap.org/resource/8084.aspx)).
- Benham, C., Cascioli Sharp, R., Chamberlain, P., Frederick, J., Gorgonio, R., Jensen, K., Rivera, J. and Veso, Y. (2014)** *Real-time evaluation of World Vision's response to typhoon Haiyan*. Monrovia, CA: World Vision. ([www.alnap.org/resource/20190.aspx](http://www.alnap.org/resource/20190.aspx)).
- Benini, A. (2009)** *Text analysis under time pressure: Tools for humanitarian and development workers*. ([www.alnap.org/resource/8189.aspx](http://www.alnap.org/resource/8189.aspx)).

**Bennett, J., Pantuliano, S., Fenton, W., Vaux, A., Barnett, C. and Brusset, E. (2010)** *Aiding the peace: A multi-donor evaluation of support to conflict prevention and peacebuilding activities in Southern Sudan 2005-2010*. Gatineau: CIDA. ([www.alnap.org/resource/8287.aspx](http://www.alnap.org/resource/8287.aspx)).

**Bentley, M. E., Boot, M., T., Gittelsohn, J. and Stallings, R. Y. (1994)** *The use of structured observations in the study of health behaviour*. The Hague: IRC International Water and Sanitation Centre. ([www.alnap.org/resource/22863.aspx](http://www.alnap.org/resource/22863.aspx)).

**Berenson, M. L., Levine, D. M. and Szabat, K. A. (2013)** 'Estimation and sample size determination for finite populations', in *Basic business statistics: global edition*. Harlow: Pearson Education Limited. ([www.alnap.org/resource/22864.aspx](http://www.alnap.org/resource/22864.aspx)).

**Berger, R., Pat-Horenczyk, R. and Gekkopf, M. (2007)** 'School-based intervention for prevention and treatment of elementary-students' terror-related distress in Israel: A quasi-randomized controlled trial'. *Journal of Traumatic Stress*, 20(4): 541-551. ([www.alnap.org/resource/22865.aspx](http://www.alnap.org/resource/22865.aspx)).

**Berk, R.A. (1983)** 'An introduction to sample selection bias in sociological data'. *American Sociological Review*, 48(3): 386-398. ([www.alnap.org/resource/22866.aspx](http://www.alnap.org/resource/22866.aspx)).

**Berlemann, M., Steinhardt, M. F. and Tutt, J. (2015)** *Do natural disasters stimulate individual saving? Evidence from a natural experiment in a highly developed country*. Bonn: Institute for the Study of Labour. ([www.alnap.org/resource/22867.aspx](http://www.alnap.org/resource/22867.aspx)).

**BetterEvaluation. (2013)** *Understand causes of outcomes and impacts*. Melbourne: BetterEvaluation. ([www.alnap.org/resource/19077.aspx](http://www.alnap.org/resource/19077.aspx)).

**Bhattacharjee, A., Jacob, P. M., Sumasundaram, M., Ramachandran, R., Thileepan, S., Kumar, S., Srikantharajah, S. and Suppiah, L. (2007)** *Final evaluation of CARE Australia supported tsunami response in Trincomalee and Batticaloa districts of Sri Lanka*. Canberra: CARE Australia. ([www.alnap.org/resource/6029.aspx](http://www.alnap.org/resource/6029.aspx)).

**Bhattacharjee, A., Sida, L. and Reddick, M. (2010)** *Evaluation of DFID-Unicef programme of cooperation: Investing in humanitarian action phase III 2006-2009*. New York: UNICEF. ([www.alnap.org/resource/5895.aspx](http://www.alnap.org/resource/5895.aspx)).

**Bhattacharjee, A., Wynter, A., Baker, J., Varghese M. and Lowery, C. (2011)** *Independent review of UNICEF's operational response to the January 2010 earthquake in Haiti*. New York: UNICEF. ([www.alnap.org/resource/6353.aspx](http://www.alnap.org/resource/6353.aspx)).

**Bilukha, O. (2008)** 'Old and new cluster designs in emergency field surveys: in search of a one-fits-all solution'. *Emerging Themes in Epidemiology*, 5(1). ([www.alnap.org/resource/8085.aspx](http://www.alnap.org/resource/8085.aspx)).

**Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., de Onis, M., Ezzati, M., Mathers, C., and Rivera, J. (2008)** 'Maternal and child undernutrition: global and regional exposures and health consequences'. *The Lancet*. ([www.alnap.org/resource/8086.aspx](http://www.alnap.org/resource/8086.aspx)).

**Blake, C. F. (1981)** 'Graffiti and racial insults: the archaeology of ethnic relations in Hawaii', in Gould R. A. and Schiffer, M. B. (eds) *Modern material culture: The archaeology of us*. New York: Academic Press. ([www.alnap.org/resource/22868.aspx](http://www.alnap.org/resource/22868.aspx)).

**Bliss, D. and Campbell, J. (2007)** *Recovering from the Java earthquake: Perceptions of the affected*. San Francisco, CA: Fritz Institute. ([www.alnap.org/resource/8087.aspx](http://www.alnap.org/resource/8087.aspx)).

**Bliss, D. and Larsen, L. (2006)** *Surviving the Pakistan earthquake: Perceptions of the affected one year later*. San Francisco, CA: Fritz Institute. ([www.alnap.org/resource/8088.aspx](http://www.alnap.org/resource/8088.aspx)).

**Boller, K., Buek, K., Burwick, A., Chatterji, M., Paulsell, D., Amin, S., Borkum, E., Campuzano, L., Jacobson, J., Sattar, S., Kadel, S., Pholy, S., Rutajwaha, A., and Sabaa, S. (2011)** *Evaluation of UNICEF's early childhood development programme with focus on government of Netherlands funding (2008-2010): Global synthesis report*. New York: UNICEF. ([www.alnap.org/resource/8089.aspx](http://www.alnap.org/resource/8089.aspx)).

**Bonbright, D. (2012)** *Use of impact evaluation results*. Washington, DC: InterAction. ([www.alnap.org/resource/8534.aspx](http://www.alnap.org/resource/8534.aspx)).

**Bond. (2013)** *Principles and checklist for assessing the quality of evidence*. London: Bond. ([www.alnap.org/resource/22807.aspx](http://www.alnap.org/resource/22807.aspx)).

**Bonino, F. (2014)** *Evaluating protection in humanitarian action : Issues and challenges*. London: ALNAP. ([www.alnap.org/resource/19237.aspx](http://www.alnap.org/resource/19237.aspx)).

**Bonino, F., Jean, I. and Knox Clarke, P. (2014)** *Closing the loop: Effective feedback in humanitarian contexts*. London: ALNAP. ([www.alnap.org/resource/10676.aspx](http://www.alnap.org/resource/10676.aspx)).

**Boonstra, E., Lindbaek, M., Fidzani, B. and Bruusgaard, D. (2001)** 'Cattle eradication and malnutrition in under fives: a natural experiment in Botswana'. *Public Health Nutrition*, 4(4): 877-882. ([www.alnap.org/resource/22869.aspx](http://www.alnap.org/resource/22869.aspx)).

**Boothby, N., Crawford, J. and Halperin, J. (2006)** 'Mozambique child soldier life outcome study: lessons learned in rehabilitation and reintegration efforts'. *Glob Public Health*, 1(1): 87–107. ([www.alnap.org/resource/22870.aspx](http://www.alnap.org/resource/22870.aspx)).

**Borton, J. (ed.) (1994)** *Code of conduct for the International Red Cross and Red Crescent movement and NGOs in disaster relief*. Geneva: IFRC. ([www.alnap.org/resource/8091.aspx](http://www.alnap.org/resource/8091.aspx)).

**Borton, J., Brusset, E. and Hallam, A. (1996)** *The international response to conflict and genocide: Lessons from the Rwanda experience: Humanitarian aid and effects (JEEAR)*. London: ODI. ([www.alnap.org/resource/2517.aspx](http://www.alnap.org/resource/2517.aspx)).

**Borton, J., Millwood, D. (1996)** *The international response to conflict and genocide : Lessons from the Rwanda experience: Study 3: Humanitarian aid and effects*. London: ODI. ([www.alnap.org/resource/9976.aspx](http://www.alnap.org/resource/9976.aspx)).

**Box, G. E. P. and Tiao, G. C. (1975)** 'Intervention analysis with applications to economic and environmental problems'. *Journal of the American Statistical Association*, 70(349): 70-79. ([www.alnap.org/resource/22871.aspx](http://www.alnap.org/resource/22871.aspx)).

**Brancati, D. (2007)** 'Political aftershocks: the impact of earthquakes on intrastate conflict'. *Journal of Conflict Resolution*, 51(5). ([www.alnap.org/resource/8092.aspx](http://www.alnap.org/resource/8092.aspx)).

**Breier, H. (2005)** *Joint evaluations: Recent experiences, lessons learned and options for the future*. Paris: OECD. ([www.alnap.org/resource/8093.aspx](http://www.alnap.org/resource/8093.aspx)).

**British Parachute Association. (2015)** How safe? ([www.alnap.org/resource/22872.aspx](http://www.alnap.org/resource/22872.aspx)).

**Brookhart, M. A. Wyss, R., Layton, J. B. and Stürmer, T. (2013)** 'Propensity score methods for confounding control in nonexperimental research'. *Circulation: Cardiovascular Quality and Outcomes*. ([www.alnap.org/resource/22873.aspx](http://www.alnap.org/resource/22873.aspx)).

- Broughton, B., Maguire, S. and David-Toweh, K. (2006)** *Inter-agency real-time evaluation of the humanitarian response to the Darfur crisis*. New York: OCHA. ([www.alnap.org/resource/3245.aspx](http://www.alnap.org/resource/3245.aspx)).
- Brown, D. and Donini, A. (2014a)** *Rhetoric or reality? Putting affected people at the centre of humanitarian action*. London: ALNAP. ([www.alnap.org/resource/12859.aspx](http://www.alnap.org/resource/12859.aspx)).
- Brown, D., Donini, A., and Knox Clarke, P. (2014b)** *Engagement of crisis-affected people in humanitarian action. ALNAP Background paper*. London: ALNAP. ([www.alnap.org/resource/10439.aspx](http://www.alnap.org/resource/10439.aspx)).
- Brown, M., Neves, J., Sandison, P., Buchanan-Smith, M., and Wiles, P. (2005)** *Evaluation of DFID-UNICEF programme of cooperation to strengthen UNICEF programming as it applies to humanitarian response, 2000-2005*. New York: UNICEF. ([www.alnap.org/resource/3355.aspx](http://www.alnap.org/resource/3355.aspx)).
- Brusset, E., Bhatt, M., Bjornestad, K., Cosgrave, J., Davies, A., Deshmukh, Y., Haleem, J., Hidalgo, S., Immajati, Y., Jayasundere, R., Mattsson, A., Muhaimin, N., Polastro, R. and Wu, T. (2009)** *A ripple in development? Long term perspectives on the response to the Indian Ocean tsunami 2004*. Stockholm: SIDA. ([www.alnap.org/resource/5679.aspx](http://www.alnap.org/resource/5679.aspx)).
- Brusset, E., Pramana, W., Davies, A., Deshmukh, Y. and Pedersen, S. (2006)** *Links between relief, rehabilitation and development in the tsunami response: Indonesia Case Study*. London: TEC. ([www.alnap.org/resource/5421.aspx](http://www.alnap.org/resource/5421.aspx)).
- Bryman, A. (2004)** 'Triangulation', in Lewis-Beck, M. S., Bryman, A. and Liao, T. F. (eds) *Encyclopedia of social science research methods*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22874.aspx](http://www.alnap.org/resource/22874.aspx)).
- Buchanan-Smith, M., Leyland, T., Aklilu, Y., Noor, M., Ahmed, S., Dini, S. and Robinson, I. (2013)** *Evaluation of FAO's cooperation in Somalia: 2007 to 2012*. Rome: FAO. ([www.alnap.org/resource/8820.aspx](http://www.alnap.org/resource/8820.aspx)).
- Buchanan-Smith, M. et al. (2005)** 'How the Sphere Project came into being: A case study of policy making in the humanitarian-aid sector and the relative influence of research', in Court, J., Hovland, I. and Young, J. (eds). *Bridging research and policy in development: Evidence and the change process*. London: ODI. ([www.alnap.org/resource/23454.aspx](http://www.alnap.org/resource/23454.aspx)).

- Buchanan-Smith, M., Ong, J. C. and Routley, S. (2015)** *Who's listening? Accountability to affected people in the Haiyan response*. Woking: Plan International. ([www.alnap.org/resource/20632.aspx](http://www.alnap.org/resource/20632.aspx)).
- Burns, M., Rowland, M., N'Guessan, R., Carneiro, I., Beeche, A., Sesler Ruiz, S., Kamara, S., Takken, W., Carnevale, P. and Allan R. (2012)** 'Insecticide-treated plastic sheeting for emergency malaria prevention and shelter among displaced populations: An observational cohort study in a refugee setting in Sierra Leone'. *The American Journal of Tropical Medicine and Hygiene*, 87(2): 242–250. ([www.alnap.org/resource/22875.aspx](http://www.alnap.org/resource/22875.aspx)).
- Bush, J. and Ati, H.A. (2007)** *Oxfam's cash transfers in the Red Sea state*. Oxford: Oxfam. ([www.alnap.org/resource/8187.aspx](http://www.alnap.org/resource/8187.aspx)).
- Bush, K. and Duggan, C. (2013)** 'Evaluation in conflict zones: Methodological and ethical challenges'. *Journal of Peacebuilding & Development*, 8(2): 5–25. ([www.alnap.org/resource/22876.aspx](http://www.alnap.org/resource/22876.aspx)).
- Buttenheim, A. (2009)** *Impact evaluations in the post-disaster setting: A conceptual discussion in the context of the 2005 Pakistan earthquake*. New Delhi: 3ie. ([www.alnap.org/resource/8206.aspx](http://www.alnap.org/resource/8206.aspx)).
- Caliendo, M. and Kopeinig, S. (2005)** *Some practical guidance for the implementation of propensity score matching*. Bonn: IZA. ([www.alnap.org/resource/22877.aspx](http://www.alnap.org/resource/22877.aspx)).
- Cameron, L. A. and Shah, M. (2012)** *Risk-taking behavior in the wake of natural disasters*. Bonn: IZA. ([www.alnap.org/resource/22878.aspx](http://www.alnap.org/resource/22878.aspx)).
- Campo, R. (2006)** "Anecdotal evidence": why narratives matter to medical practice'. *PLoS Medicine*, 3(10), 1677-1678. ([www.alnap.org/resource/22879.aspx](http://www.alnap.org/resource/22879.aspx)).
- Canteli, C., Morris, T. and Steen, N. (2012)** *Synthesis of mixed method impact evaluations of the contribution of food assistance to durable solutions in protracted refugee situations*. Rome: WFP. ([www.alnap.org/resource/19902.aspx](http://www.alnap.org/resource/19902.aspx)).
- Carden, F. (2009)** *Knowledge to policy: making the most of development research*. New Delhi: SAGE. ([www.alnap.org/resource/22881.aspx](http://www.alnap.org/resource/22881.aspx)).
- Carlsson, J., Eriksson-Baaz, M., Fallenius, A. and Lovgren, E. (1999)** *Are evaluations useful? Cases from Swedish Development co-operation*. Stockholm: SIDA. ([www.alnap.org/resource/12371.aspx](http://www.alnap.org/resource/12371.aspx)).



- Catalano, R., Bruckner, T., Gould, J., Eskenazi, B., and Anderson, E. (2005)** 'Sex ratios in California following the terrorist attacks of September 11, 2001'. *Human Reproduction*, 20(5): 1221–1227. ([www.alnap.org/resource/22882.aspx](http://www.alnap.org/resource/22882.aspx)).
- CRS. (2012)** *CRS Haiti accountability framework*. Baltimore, MD: CRS. ([www.alnap.org/resource/10549.aspx](http://www.alnap.org/resource/10549.aspx)).
- Catley, A., Burns, J., Abebe, D. and Suji, O. (2008)** *Participatory impact assessment: A guide for practitioners*. Somerville, MA: Feinstein International Centre. ([www.alnap.org/resource/8094.aspx](http://www.alnap.org/resource/8094.aspx)).
- Catley, A., Burns, J., Abebe, D. and Suji, O. (2013)** *Participatory impact assessment: A design guide*. Somerville, MA: Feinstein International Centre. ([www.alnap.org/resource/10811.aspx](http://www.alnap.org/resource/10811.aspx)).
- CDC. (2009)** *Data collection methods for program evaluation: interviews*. Atlanta, GA: CDC. ([www.alnap.org/resource/19260.aspx](http://www.alnap.org/resource/19260.aspx)).
- Channel Research. (2011)** *5-year evaluation of the Central Emergency Response Fund*. London: Channel Research. ([www.alnap.org/resource/6357.aspx](http://www.alnap.org/resource/6357.aspx)).
- Chapman, N. and Vaillant, C. (2010)** *Synthesis of country programme evaluations conducted in fragile states*. London: DFID. ([www.alnap.org/resource/6358.aspx](http://www.alnap.org/resource/6358.aspx)).
- Chatham House. (2007)** *Chatham house rules*. London: Chatham House. ([www.alnap.org/resource/8097.aspx](http://www.alnap.org/resource/8097.aspx)).
- Checchi, F., Roberts, B. and Morgan, O. (2009)** *A new method to estimate mortality in crisis-affected populations: Validation and feasibility study*. Washington, DC: USAID. ([www.alnap.org/resource/22883.aspx](http://www.alnap.org/resource/22883.aspx)).
- Christensen, G. (2016)** *Manual of best practices in transparent social science research*. Berkeley, CA: Berkeley Initiative for Transparency in the Social Sciences. ([www.alnap.org/resource/22884.aspx](http://www.alnap.org/resource/22884.aspx)).
- Christoplos, I. (2006)** *Links between relief, rehabilitation and development in the tsunami response*. London: TEC. ([www.alnap.org/resource/3533.aspx](http://www.alnap.org/resource/3533.aspx)).
- CHS Alliance. (2014)** *Core humanitarian standard on quality and accountability*. Geneva: CHS Alliance. ([www.alnap.org/resource/22799.aspx](http://www.alnap.org/resource/22799.aspx)).

- CHS Alliance. (2015)** *Core Humanitarian CHS Guidance Notes and Indicators*. Geneva: CHS Alliance. ([www.alnap.org/resource/22885.aspx](http://www.alnap.org/resource/22885.aspx)).
- CIDA. (2000)** *How to perform evaluations: Model ToR*. Gatineau: CIDA. ([www.alnap.org/resource/22886.aspx](http://www.alnap.org/resource/22886.aspx)).
- CIDA. (2005)** *Additional resources: Accompanying document of the environment handbook for community development initiatives*. Gatineau: CIDA. ([www.alnap.org/resource/8098.aspx](http://www.alnap.org/resource/8098.aspx)).
- Clarke, N., Loveless, J., Ojok, B., Routley, S. and Vaux, T. (2015)** *Report of the inter-agency humanitarian evaluation (IAHE) of the response to conflict in South Sudan*. New York: OCHA. ([www.alnap.org/resource/22828.aspx](http://www.alnap.org/resource/22828.aspx)).
- Collier, D. and Mahoney, J. (1996)** 'Insights and pitfalls: Selection bias in qualitative research'. *World Politics*, 49(1): 56–91. ([www.alnap.org/resource/22887.aspx](http://www.alnap.org/resource/22887.aspx)).
- Collinson, S. and Elhawary, S. (2012)** *Humanitarian space: a review of trends and issues*. London: ODI. ([www.alnap.org/resource/22888.aspx](http://www.alnap.org/resource/22888.aspx)).
- Cook, T.D., Scriven, M., Coryn, C. L. S. and Evergreen, S. D. H. (2010)** 'Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven'. *American Journal of Evaluation*, 31(1): 105–117. ([www.alnap.org/resource/22323.aspx](http://www.alnap.org/resource/22323.aspx)).
- Corlazzoli, V. and White, J. (2013)** *Practical approaches to theories of change in conflict, security, and justice. Part II: Using theories of change in monitoring and evaluation*. London: DFID. ([www.alnap.org/resource/22890.aspx](http://www.alnap.org/resource/22890.aspx)).
- Cornwall, A. (2014)** *Using participatory process evaluation to understand the dynamics of change in a nutrition education programme*. Brighton: IDS. ([www.alnap.org/resource/22891.aspx](http://www.alnap.org/resource/22891.aspx)).
- Cosgrave, J. (2004)** *Danish assistance to internally displaced persons in Angola 1999-2003*. London: Channel Research. ([www.alnap.org/resource/22892.aspx](http://www.alnap.org/resource/22892.aspx)).
- Cosgrave, J. (2008)** *Responding to earthquakes 2008: Learning from earthquake relief and recovery operations*. London: ALNAP. ([www.alnap.org/resource/5239.aspx](http://www.alnap.org/resource/5239.aspx)).

- Cosgrave, J. (2010)** *Evaluability assessment of a proposed evaluation of humanitarian interventions in South and Central Somalia: Commissioned by the IASC for Somalia: overall report*. Geneva: IASC. ([www.alnap.org/resource/22893.aspx](http://www.alnap.org/resource/22893.aspx)).
- Cosgrave, J., Crawford, N. and Mosel, I. (2015)** *10 Things to know about refugees and displacement*. London: HPG/ODI. ([www.alnap.org/resource/22895.aspx](http://www.alnap.org/resource/22895.aspx)).
- Cosgrave, J., Goncalves, C., Martyris, D., Polastro, R. and Sikumba-Dils, M. (2007)** *Inter-agency real-time evaluation of the response to the February 2007 floods and cyclone in Mozambique*. Madrid: DARA. ([www.alnap.org/resource/3457.aspx](http://www.alnap.org/resource/3457.aspx)).
- Cosgrave, J., Polastro, R. and Zafar, F. (2010)** *Inter-agency real time evaluation (IA RTE) of the humanitarian response to Pakistan's 2009 displacement crisis*. Geneva: IASC. ([www.alnap.org/resource/11896.aspx](http://www.alnap.org/resource/11896.aspx)).
- Cosgrave, J., Ramalingam, B. and Beck, T. (2009)** *Real-time evaluations of humanitarian action*. ALNAP Guide. London: ALNAP. ([www.alnap.org/resource/5595.aspx](http://www.alnap.org/resource/5595.aspx)).
- Cosgrave, J., Selvester, K., Fidalgo, L., Hallam, A. and Taimo, N. (2001)** *Independent evaluation of DEC Mozambique floods appeal funds March 2000 - December 2000 : Volume One: Main Findings*. London: DEC. ([www.alnap.org/resource/2999.aspx](http://www.alnap.org/resource/2999.aspx)).
- Cosgrave, J., Wata, H., Ntata, P., Immajati, Y. And Bhatt, M. (2010)** *Programme evaluation of disaster risk reduction commissioned by Cordaid overall report*. London: Channel Research. ([www.alnap.org/resource/12523.aspx](http://www.alnap.org/resource/12523.aspx)).
- Cossée, O., Belli, L., Bultemeier, B. and Carrugi, C. (2010)** *Evaluation of FAO interventions funded by the CERF: Final report*. Rome: FAO. ([www.alnap.org/resource/8099.aspx](http://www.alnap.org/resource/8099.aspx)).
- Court, J. and Young, J. (2003)** *Bridging research and policy: Insights from 50 case studies*. London: ODI. ([www.alnap.org/resource/8100.aspx](http://www.alnap.org/resource/8100.aspx)).
- CP MERG. (2012)** *Ethical principles, dilemmas and risks in collecting data on violence against children: A review of available literature*. CP MERG. ([www.alnap.org/resource/22327.aspx](http://www.alnap.org/resource/22327.aspx)).

- Crawford, N., Cosgrave, J., Haysom, S., Walicki, N. (2015)** *Protracted displacement: uncertain paths to self-reliance in exile*. London: HPG/ODI. ([www.alnap.org/resource/21304.aspx](http://www.alnap.org/resource/21304.aspx)).
- Crawford, P., Bysouth, K., Nichols, D. and Thompon, F. (2006)** *CAER cluster evaluation: Pakistan earthquake*. Canberra: AusAID. ([www.alnap.org/resource/8190.aspx](http://www.alnap.org/resource/8190.aspx)).
- Crawford, P. and Eagles, J. (2007)** *ANCP Pacific cluster evaluation report*. Version 2.0. Canberra: AusAID. ([www.alnap.org/resource/8191.aspx](http://www.alnap.org/resource/8191.aspx)).
- Creswell, J.W. (2002)** *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. 4th ed. Upper Saddle River: Pearson. ([www.alnap.org/resource/22896.aspx](http://www.alnap.org/resource/22896.aspx)).
- Creswell, J. W. and Plano Clark, V. L. (2011)** *Designing and conducting mixed methods research*. 2nd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22897.aspx](http://www.alnap.org/resource/22897.aspx)).
- Crewe, E. and Young, J. (2002)** *Bridging research and policy: Context, evidence and links*. London: ODI. ([www.alnap.org/resource/22898.aspx](http://www.alnap.org/resource/22898.aspx)).
- Crisp, J., Garras, G., McAvoy, J., Schenkenberg, E., Speigel, P. and Voon, F. (2013)** *From slow boil to breaking point: A real-time evaluation of UNHCR's response to the Syrian refugee emergency*. Geneva: UNHCR. ([www.alnap.org/resource/8848.aspx](http://www.alnap.org/resource/8848.aspx)).
- Curtis, V. and Kanki, B. (1999)** *A manual on hygiene promotion*. New York: UNICEF. ([www.alnap.org/resource/22899.aspx](http://www.alnap.org/resource/22899.aspx)).
- Dabelstein, N. (1996)** 'Evaluating the international humanitarian system: rationale, process and management of the joint evaluation of the international response to the Rwanda genocide. *Disasters*, 20(4). ([www.alnap.org/resource/8101.aspx](http://www.alnap.org/resource/8101.aspx)).
- Dahlgren, A.-L., DeRoo, L. A., Avril, J. and Loutan, L. (2009)** 'Health risks and risk-taking behaviors among International Committee of the Red Cross (ICRC) expatriates returning from humanitarian missions. *Journal of Travel Medicine*, 16(6): 382–390. ([www.alnap.org/resource/22900.aspx](http://www.alnap.org/resource/22900.aspx)).

**Daito, H., Suzuki, M., Shiihara, J., Kilgore, P. E., Ohtomo, H., Morimoto, K., Ishida, M., Kamigaki, T., Oshitani, H., Hashizume, M., Endo, W., Hagiwara, K., Ariyoshi, K. and Okinaga, S. (2013)** 'Impact of the Tohoku earthquake and tsunami on pneumonia hospitalisations and mortality among adults in northern Miyagi, Japan: a multicentre observational study'. *Thorax*, 68(6): 544–550. ([www.alnap.org/resource/22901.aspx](http://www.alnap.org/resource/22901.aspx)).

**DANIDA. (2012)** *Evaluation of Danish development support to Afghanistan*. Copenhagen: DANIDA. ([www.alnap.org/resource/22903.aspx](http://www.alnap.org/resource/22903.aspx)).

**Daniel, J., (2013)** *Sampling essentials: Practical guidelines for making sampling choices*. Los Angeles, CA: SAGE. ([www.alnap.org/resource/9161.aspx](http://www.alnap.org/resource/9161.aspx)).

**Das, N. C. and Shams, R. (2011)** *Asset transfer programme for the ultra poor: A randomized control trial evaluation*. Dhaka: BRAC. ([www.alnap.org/resource/22904.aspx](http://www.alnap.org/resource/22904.aspx)).

**Davidson, E. J. (2007)** 'Unlearning some of our social scientist habits'. *Journal of MultiDisciplinary Evaluation*, 4(8): iii–vi. ([www.alnap.org/resource/22906.aspx](http://www.alnap.org/resource/22906.aspx)).

**Davidson, E. J. (2009)** *Causal inference: Nuts and bolts: A mini workshop for the anzea Wellington branch*. Wellington: Davidson Consulting. ([www.alnap.org/resource/22908.aspx](http://www.alnap.org/resource/22908.aspx)).

**Davidson, E. J. (2010a)** '9 golden rules for commissioning a waste-of-money evaluation', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22909.aspx](http://www.alnap.org/resource/22909.aspx)).

**Davidson, E. J. (2010b)** 'Commissioning XGEMs – the sequel', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22910.aspx](http://www.alnap.org/resource/22910.aspx)).

**Davidson, E. J. (2010c)** 'Extreme genuine evaluation makeovers (XGEMs) for commissioning', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22912.aspx](http://www.alnap.org/resource/22912.aspx)).

**Davidson, E. J. (2012a)** '6 questions that cut to the chase when choosing the right evaluation team', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22905.aspx](http://www.alnap.org/resource/22905.aspx)).

**Davidson, E. J. (2012b)** '9 hot tips for commissioning, managing (and doing!) actionable evaluation' in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22913.aspx](http://www.alnap.org/resource/22913.aspx)).

**Davidson, E. J. (2014a)** *Evaluative Reasoning*. Florence: UNICEF. ([www.alnap.org/resource/22331.aspx](http://www.alnap.org/resource/22331.aspx)).

**Davidson, E. J. (2014b)** "'Minirubrics' – 7 hot tips for using this cool tool to focus evaluative conversations', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22914.aspx](http://www.alnap.org/resource/22914.aspx)).

**Davies, R. and Dart, J. (2005)** *The "most significant change" (MSC) technique: A guide to its use*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8102.aspx](http://www.alnap.org/resource/8102.aspx)).

**Davin, E., Gonzalez, V. and Majidi, N. (2009)** *UNHCR's voluntary repatriation program: Evaluation of the impact of the cash grant*. Geneva: UNHCR. ([www.alnap.org/resource/22915.aspx](http://www.alnap.org/resource/22915.aspx)).

**Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H. and Resnicow, K. (2010)** 'Interviewer effects in public health surveys'. *Health Education Research*, 25(1): 14–26. ([www.alnap.org/resource/8103.aspx](http://www.alnap.org/resource/8103.aspx)).

**Day, S. J. and Altman, D. G. (2000)** 'Blinding in clinical trials and other studies'. *BMJ*, 321(8): 504. ([www.alnap.org/resource/22917.aspx](http://www.alnap.org/resource/22917.aspx)).

**DEC. (2013)** *DEC Approach to Learning*. London: DEC. ([www.alnap.org/resource/8649.aspx](http://www.alnap.org/resource/8649.aspx)).

**de Nicola, F. and Giné, X. (2011)** *How accurate are recall data? Evidence from coastal India*. Washington, DC: World Bank. ([www.alnap.org/resource/8104.aspx](http://www.alnap.org/resource/8104.aspx)).

**Denzin, N. K. (2010)** 'Moments, mixed methods, and paradigm dialogs'. *Qualitative Inquiry*, 16(6): 419–427. ([www.alnap.org/resource/8105.aspx](http://www.alnap.org/resource/8105.aspx)).

**de Ville de Goyet, C. and Morinière, L. (2006)** *The role of needs assessment in the tsunami response*. London: TEC. ([www.alnap.org/resource/12406.aspx](http://www.alnap.org/resource/12406.aspx)).

**de Ville de Goyet, C., Scheuermann, P., Reed, S. B., Al Wehaidy, R., and Termil, A. (2011)** *SDC humanitarian aid: Emergency relief*. Bern: Swiss Agency for Development and Cooperation. ([www.alnap.org/resource/6142.aspx](http://www.alnap.org/resource/6142.aspx)).

**de Ville de Goyet, C., Buckle, P., Ievers, J., Parlan, H. P. and Viandrito, J. (2013)** *Evaluation of the European Commission's humanitarian and disaster risk reduction activities (DIPECHO) in Indonesia*. Brussels: ECHO. ([www.alnap.org/resource/11739.aspx](http://www.alnap.org/resource/11739.aspx)).

**DFID. (2003)** *Tools for development: A handbook for those engaged in development activity*. London: DFID. ([www.alnap.org/resource/11859.aspx](http://www.alnap.org/resource/11859.aspx)).

**DFID, (2005)** *Guidance and evaluation and review for DFID staff*. London: DFID. ([www.alnap.org/resource/22918.aspx](http://www.alnap.org/resource/22918.aspx)).

**DFID. (2010)** *Evaluation study terms of reference (template)*. London: DFID. ([www.alnap.org/resource/22920.aspx](http://www.alnap.org/resource/22920.aspx)).

**DFID. (2014)** *Assessing the strength of evidence*. London: DFID. ([www.alnap.org/resource/10949.aspx](http://www.alnap.org/resource/10949.aspx)).

**DFID and UKAID. (2011)** *Multilateral aid review: ensuring maximum value for money for UK aid through multilateral organisations*. London: DFID and UKAID. ([www.alnap.org/resource/8106.aspx](http://www.alnap.org/resource/8106.aspx)).

**Dillman, D. A., Smyth, J. D. and Christian, L. M. (2009)** *Internet, mail and mixed-mode surveys: The tailored design method*. 3rd ed. Hoboken, NJ: Wiley. ([www.alnap.org/resource/8107.aspx](http://www.alnap.org/resource/8107.aspx)).

**Disch, A., Haarberg, K., Gharney, A. B. and Lunøe, B. (2008)** *Synthesis study on best practices and innovative approaches to capacity development in low-income African countries*. Oslo: NORAD. ([www.alnap.org/resource/10132.aspx](http://www.alnap.org/resource/10132.aspx)).

**Dohrenwend, B. S., Colombotos, J. and Dohrenwend, B. P. (1968)** 'Social distance and interviewer effects'. *Public Opinion Quarterly*, 32(3): 410–422. ([www.alnap.org/resource/8109.aspx](http://www.alnap.org/resource/8109.aspx)).

**Drucker, P. F. (1985)** *Innovation and entrepreneurship*. Oxford: Butterworth-Heinemann. ([www.alnap.org/resource/22921.aspx](http://www.alnap.org/resource/22921.aspx)).

**Dryden-Peterson, S. (2011)** *Refugee education: A global review*. Geneva: UNHCR. ([www.alnap.org/resource/8110.aspx](http://www.alnap.org/resource/8110.aspx)).

**Duncalf, J. (2013)** *A real-time evaluation of ACF International's response to typhoon Haiyan / Yolanda in the Philippines*. Paris: ACF. ([www.alnap.org/resource/20994.aspx](http://www.alnap.org/resource/20994.aspx)).

**du Preez, M., Conroy, R. M., Wright, J. A., Moyo, S., Potgieter, N. and Gundry, S. W. (2008)** 'Use of ceramic water filtration in the prevention of diarrheal disease: A randomized controlled trial in rural South Africa and Zimbabwe. *The American Journal of Tropical Medicine and Hygiene*, 79(5): 696–701. ([www.alnap.org/resource/22922.aspx](http://www.alnap.org/resource/22922.aspx)).

**Durkin, M. S., Khan, N., Davidson, L. L., Zaman, S. S. and Stein, Z. A. (1993)** 'The effects of a natural disaster on child behavior: evidence for posttraumatic stress'. *American Journal of Public Health*, 83(11): 1549–1553. ([www.alnap.org/resource/12530.aspx](http://www.alnap.org/resource/12530.aspx)).

**Ebbinghaus, B. (2005)** 'When less is more: Selection problems in large- N and small- N cross-national comparisons. *International Sociology*, 20(2): 133–152. ([www.alnap.org/resource/22924.aspx](http://www.alnap.org/resource/22924.aspx)).

**ECHO. (2002)** *Evaluating humanitarian action funded by the humanitarian aid office of the European Commission*. Brussels: ECHO. ([www.alnap.org/resource/22925.aspx](http://www.alnap.org/resource/22925.aspx)).

**Edmondson, A. C. (2004)** 'Learning from failure in health care: frequent opportunities, pervasive barriers'. *Quality and Safety in Health Care*, 13(2):ii3–ii9. ([www.alnap.org/resource/8111.aspx](http://www.alnap.org/resource/8111.aspx)).

**ECB. (2011a)** *What we know about joint evaluations of humanitarian action*. London: ECB. ([www.alnap.org/resource/19286.aspx](http://www.alnap.org/resource/19286.aspx)).

**ECB. (2011b)** *What we know about joint evaluations of humanitarian action. Section 1 of 3: the guide*. London: ECB. ([www.alnap.org/resource/9536.aspx](http://www.alnap.org/resource/9536.aspx)).

**ECB. (2011c)** *What we know about joint evaluations of humanitarian action. Section 2 of 3: the stories*. London: ECB. ([www.alnap.org/resource/9537.aspx](http://www.alnap.org/resource/9537.aspx)).

**ECB. (2011d)** *What we know about joint evaluations of humanitarian action. Section 3 of 3: the tools*. London: ECB. ([www.alnap.org/resource/9538.aspx](http://www.alnap.org/resource/9538.aspx)).

**ECB. (2012)** *Joint evaluations. Guidelines from the ECB project*. London: ECB. ([www.alnap.org/resource/19413.aspx](http://www.alnap.org/resource/19413.aspx)).

**ECB. (2013)** *What we know about collaboration: the ECB country consortium experience*. London: ECB. ([www.alnap.org/resource/10552.aspx](http://www.alnap.org/resource/10552.aspx)).



**EES. (2007)** *EES statement: The importance of a methodologically diverse approach to impact evaluation - specifically with respect to development aid and development interventions*. Prague: EES. ([www.alnap.org/resource/11119.aspx](http://www.alnap.org/resource/11119.aspx)).

**EES. (2015)** *The EES capabilities framework*. Prague: EES. ([www.alnap.org/resource/22927.aspx](http://www.alnap.org/resource/22927.aspx)).

**EPOC. (2013)** *Interrupted time series (ITS) analyses*. Oslo: Norwegian Knowledge Centre for the Health Services. ([www.alnap.org/resource/22928.aspx](http://www.alnap.org/resource/22928.aspx)).

**Egeland, J., Harmer, A. and Stoddard, A. (2011)** *To stay and deliver: Good practice for humanitarian in complex security environments*. New York: OCHA. ([www.alnap.org/resource/6364.aspx](http://www.alnap.org/resource/6364.aspx)).

**Eisinga, R., te Grotenhuis, M., Larsen, J. K., Pelzer, B. and van Strien, T. (2011)** BMI of interviewer effects. *International Journal of Public Opinion Research*, 23(4): 530–543. ([www.alnap.org/resource/8112.aspx](http://www.alnap.org/resource/8112.aspx)).

**FDFA. (2014)** *Humanitarian access in situations of armed conflict*. Bern: FDFA. ([www.alnap.org/resource/21733.aspx](http://www.alnap.org/resource/21733.aspx)).

**Fearon, J., Humphreys, M. and Weinstein, J. (2008)** *Community driven reconstruction in Lofa County: Impact assessment*. New York: IRC. ([www.alnap.org/resource/8192.aspx](http://www.alnap.org/resource/8192.aspx)).

**Featherstone, A. (2012)** *Evaluation of Somalia drought response 2011/2012: Using Oxfam GB's global humanitarian indicator tool*. Oxford: Oxfam. ([www.alnap.org/resource/7983.aspx](http://www.alnap.org/resource/7983.aspx)).

**Featherstone, A. (2013a)** *Improving impact: do accountability mechanisms deliver results?* Geneva: HAP. ([www.alnap.org/resource/8388.aspx](http://www.alnap.org/resource/8388.aspx)).

**Featherstone, A. (2013b)** *Syria needs analysis project (SNAP): External mid-term review*. Geneva: ACAPS. ([www.alnap.org/resource/10390.aspx](http://www.alnap.org/resource/10390.aspx)).

**Featherstone, A. (2014)** *A review of UNHCR's utilisation of the Central Emergency Response Fund*. Geneva: UNHCR. ([www.alnap.org/resource/12430.aspx](http://www.alnap.org/resource/12430.aspx)).

**Ferf, A. and Fabbri, P. (2014)** *Impact evaluation of Swiss solidarity Asian tsunami programme*. Geneva: Swiss Solidarity. ([www.alnap.org/resource/22473.aspx](http://www.alnap.org/resource/22473.aspx)).

- Fergusson, D., Cranley Glass, K., Waring, D., Shapiro, S. (2004)** 'Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials'. *BMJ*. ([www.alnap.org/resource/22930.aspx](http://www.alnap.org/resource/22930.aspx)).
- Few, R., McAvoy, D., Tarazona, M. and Walden, V. (2014)** *Contribution to Change: An approach to evaluating the role of intervention in disaster recovery*. Oxford: Oxfam. ([www.alnap.org/resource/10305.aspx](http://www.alnap.org/resource/10305.aspx)).
- Flaherty, E. W., Barry, E. and Swift, M. (1978)** 'Use of an unobtrusive measure for the evaluation of interagency coordination. *Evaluation Review*, 2(2): 261–273. ([www.alnap.org/resource/22931.aspx](http://www.alnap.org/resource/22931.aspx)).
- Flores-Macias, F. and Lawson, C. (2008)** 'Effects of interviewer gender on survey responses: Findings from a household survey in Mexico'. *International Journal of Public Opinion Research*, 20(1): 100–110. ([www.alnap.org/resource/8113.aspx](http://www.alnap.org/resource/8113.aspx)).
- Foley, P. (2009)** *Participatory evaluation of the 2008 farmer field school programme, Lira, Uganda*. Kampala: ACF. ([www.alnap.org/resource/20908.aspx](http://www.alnap.org/resource/20908.aspx)).
- Fortune, V. and Rasal, P. (2010)** *British Red Cross - mass sanitation module - 2010 Haiti earthquake response - post deployment learning evaluation*. London: British Red Cross. ([www.alnap.org/resource/6038.aspx](http://www.alnap.org/resource/6038.aspx)).
- Foubert, V. and Eskandar, H. (2009)** *Taking the initiative – Exploring quality and accountability in the humanitarian sector: an introduction to eight initiatives*. Geneva: Sphere Project. ([www.alnap.org/resource/5673.aspx](http://www.alnap.org/resource/5673.aspx)).
- Fowler, F. J. (2009)** *Survey research methods*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8115.aspx](http://www.alnap.org/resource/8115.aspx)).
- Fritz Institute. (2005)** *Recipient perceptions of aid effectiveness: rescue, relief and rehabilitation in tsunami affected Indonesia, India and Sri Lanka*. San Francisco: CA: Fritz Institute. ([www.alnap.org/resource/5459.aspx](http://www.alnap.org/resource/5459.aspx)).
- Fritz Institute. (2006)** *Hurricane Katrina: Perceptions of the affected*. San Francisco, CA: Fritz Institute. ([www.alnap.org/resource/8116.aspx](http://www.alnap.org/resource/8116.aspx)).
- Frongillo, E. A. (2012)** *Analysis of complex surveys*. Cornell: Cornell Statistical Consulting Unit. ([www.alnap.org/resource/22932.aspx](http://www.alnap.org/resource/22932.aspx)).

- Funnell, S. C. and Rogers, P. (2011)** *Purposeful program theory: Effective use of theories of change and logic models*. Hoboken, NJ: Wiley. ([www.alnap.org/resource/8193.aspx](http://www.alnap.org/resource/8193.aspx)).
- Gamarra, T., Reed, S. B. and Wilding, J. (2005)** *Final evaluation of hurricanes operation 2004*. Geneva: IFRC. ([www.alnap.org/resource/12536.aspx](http://www.alnap.org/resource/12536.aspx)).
- Garvin, D. A., Edmondson, A. C. and Gino, F. (2008)** 'Is yours a learning organization?' *Harvard Business Review*, 86(3). ([www.alnap.org/resource/8117.aspx](http://www.alnap.org/resource/8117.aspx)).
- Gerring, J. (2006)** 'Techniques for choosing cases', in *Case study research: principles and practices*. New York: Cambridge University Press. ([www.alnap.org/resource/9014.aspx](http://www.alnap.org/resource/9014.aspx)).
- Gilmour, S., Degenhardt, L., Hall, W. and Day, C. (2006)** 'Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: a general method and example'. *BMC Medical Research Methodology*, 6(1): 1–9. ([www.alnap.org/resource/22933.aspx](http://www.alnap.org/resource/22933.aspx)).
- Given, L. M. (2008)** *The SAGE encyclopedia of qualitative research methods*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22934.aspx](http://www.alnap.org/resource/22934.aspx)).
- Glass, G. V. (1997)** *Interrupted time series quasi-experiments*. Tempe, AZ: Arizona State University. ([www.alnap.org/resource/22935.aspx](http://www.alnap.org/resource/22935.aspx)).
- Glouberman, S. and Zimmerman, B. (2002)** *Complicated and complex systems: what would successful reform of Medicare look like?* Ottawa: Commission on the Future of Health Care in Canada. ([www.alnap.org/resource/8119.aspx](http://www.alnap.org/resource/8119.aspx)).
- Godden, B. (2004)** *Sample size formulas*. Chicago: Appraisal institute. ([www.alnap.org/resource/22936.aspx](http://www.alnap.org/resource/22936.aspx)).
- Goertz, G. (2012)** *A tale of two cultures: Qualitative and quantitative research in the social sciences*. Princeton: Princeton University Press. ([www.alnap.org/resource/13029.aspx](http://www.alnap.org/resource/13029.aspx)).
- Goertz, G. and Mahoney, J. (2012)** 'Methodological Rorschach tests: Contrasting interpretations in qualitative and quantitative research'. *Comparative Political Studies*, 46(2): 236–251. ([www.alnap.org/resource/22937.aspx](http://www.alnap.org/resource/22937.aspx)).

- GHD. (2003a.)** 'Good practices in donor financing, management and accountability', in *23 principles and good practice of humanitarian donorship*. ([www.alnap.org/resource/10184.aspx](http://www.alnap.org/resource/10184.aspx)).
- GHD. (2003b)** *International meeting on good humanitarian donorship*. Stockholm: GHD. ([www.alnap.org/resource/22940.aspx](http://www.alnap.org/resource/22940.aspx)).
- Goodrick, D. (2014)** *Comparative case studies*. Florence: UNICEF. ([www.alnap.org/resource/22941.aspx](http://www.alnap.org/resource/22941.aspx)).
- Gover, A. R., Macdonald, J. M. and Alpert, G. P. (2003)** 'Combating domestic violence: Findings from an evaluation of a local domestic violence court'. *Criminology & Public Policy*, 3(1): 109–132. ([www.alnap.org/resource/22942.aspx](http://www.alnap.org/resource/22942.aspx)).
- Goyder, H. (2010)** *Evaluation of DEC-funded shelter projects following the 2009 Indonesian earthquake*. Teddington: Tearfund. ([www.alnap.org/resource/5958.aspx](http://www.alnap.org/resource/5958.aspx)).
- Graig, E. (2011)** *Hiring an evaluation consultant*. Riverdale: Usable Knowledge. ([www.alnap.org/resource/22944.aspx](http://www.alnap.org/resource/22944.aspx)).
- Graig, E. (2013)** 'An alternative to the evaluation RFP process', in Usable Knowledge. ([www.alnap.org/resource/22943.aspx](http://www.alnap.org/resource/22943.aspx)).
- Grais, R. F., Luquero, F. J., Grellety, E., Pham, H., Coghlan, B. and Salignon, P. (2009)** 'Learning lessons from field surveys in humanitarian contexts: a case study of field surveys conducted in North Kivu, DRC 2006-2008'. *Conflict and Health*, 3(8). ([www.alnap.org/resource/22946.aspx](http://www.alnap.org/resource/22946.aspx)).
- Grasso, P. G. Imas, L. M. and Fostvedt, N. (2013)** *Use of evaluation in the Norwegian development cooperation system*. Oslo: Norad. ([www.alnap.org/resource/22947.aspx](http://www.alnap.org/resource/22947.aspx)).
- Gray, D.E. (2014)** *Doing research in the real world*. 3rd ed. Los Angeles: SAGE. ([www.alnap.org/resource/22948.aspx](http://www.alnap.org/resource/22948.aspx)).
- Green, D. (2012)** *Creating killer facts and graphics*. Oxford: Oxfam. ([www.alnap.org/resource/22949.aspx](http://www.alnap.org/resource/22949.aspx)).
- Groupe URD. (2009)** *Quality Compas Companion Book*. Plaisians: Groupe URD. ([www.alnap.org/resource/8541.aspx](http://www.alnap.org/resource/8541.aspx)).

- Grünewald, F., Binder, A. and Georges, Y. (2010)** *Inter-agency real-time evaluation in Haiti: 3 months after the earthquake*. Plaisians: Groupe URD. ([www.alnap.org/resource/9962.aspx](http://www.alnap.org/resource/9962.aspx)).
- Grünewald, F., Kauffmann, D., Boyer, B. and Patinet, J. (2011)** *Real-time evaluation of humanitarian action supported by DG ECHO in Haiti: 2009-2011: November 2010-April 2011*. Plaisians: Groupe URD. ([www.alnap.org/resource/22950.aspx](http://www.alnap.org/resource/22950.aspx)).
- Gubbels, P. and Bousquet, C. (2013)** *Independent evaluation of CARE's response to the 2011-2012 Sahel humanitarian crisis*. Geneva: CARE International. ([www.alnap.org/resource/20768.aspx](http://www.alnap.org/resource/20768.aspx)).
- Guerrero, S., Woodhead, S. and Hounjet, M. (2013)** *On the right track: a brief review of monitoring and evaluation in the humanitarian sector*. ([www.alnap.org/resource/8211.aspx](http://www.alnap.org/resource/8211.aspx)).
- Guest, G., Bunce, A. and Johnson, L. (2006)** How many interviews are enough? *Field methods*, 18(1): 59–82. ([www.alnap.org/resource/22952.aspx](http://www.alnap.org/resource/22952.aspx)).
- Guijt, I. (2014)** *Participatory Approaches*, Florence: UNICEF. ([www.alnap.org/resource/22798.aspx](http://www.alnap.org/resource/22798.aspx)).
- Hagens, C. and Ishida, L. (2010)** *Real time evaluation of CRS's flood response in Pakistan: KPK and Baluchistan*. Baltimore, MD: CRS. ([www.alnap.org/resource/5956.aspx](http://www.alnap.org/resource/5956.aspx)).
- Hahn, J., Todd, P. and Van der Klaauw, W. (2001)** 'Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1): 201–209. ([www.alnap.org/resource/22954.aspx](http://www.alnap.org/resource/22954.aspx)).
- Hallam, A. (2011)** *Harnessing the power of evaluation in humanitarian action: An initiative to improve understanding and use of evaluation*. London: ALNAP. ([www.alnap.org/resource/6123.aspx](http://www.alnap.org/resource/6123.aspx)).
- Hallam, A. and Bonino, F. (2013)** *Using evaluation for a change: Insights from humanitarian practitioners*. London: ALNAP. ([www.alnap.org/resource/8980.aspx](http://www.alnap.org/resource/8980.aspx)).
- Hammersley, M. (1992)** *What's wrong with ethnography?* London: Routledge. ([www.alnap.org/resource/9017.aspx](http://www.alnap.org/resource/9017.aspx)).

- Hanley, T., Binas, R., Murray, J. and Tribunalo, B. (2014)** *IASC Inter-agency humanitarian evaluation of the typhoon Haiyan response*. Geneva: IASC. ([www.alnap.org/resource/19318.aspx](http://www.alnap.org/resource/19318.aspx)).
- HAP. (2010)** *The 2010 HAP standard in accountability and quality management*. Geneva: HAP. ([www.alnap.org/resource/8125.aspx](http://www.alnap.org/resource/8125.aspx)).
- Hart, C. (1998)** *Doing a literature review: releasing the social science research imagination*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8126.aspx](http://www.alnap.org/resource/8126.aspx)).
- Hassan, P., Freese, A. and Guallar, M. (2013)** *Desk review of UNDAFs commencing in 2013*. New York: UNDP. ([www.alnap.org/resource/22811.aspx](http://www.alnap.org/resource/22811.aspx)).
- Haver, K., Hatungimana, F. and Tennant, V. (2009)** *Money matters: An evaluation of the use of cash grants in UNHCR's voluntary repatriation and reintegration programme in Burundi*. Geneva: UNHCR. ([www.alnap.org/resource/5710.aspx](http://www.alnap.org/resource/5710.aspx)).
- HelpAge International. (2010)** *A study of humanitarian funding for older people*. London: HelpAge International. ([www.alnap.org/resource/8127.aspx](http://www.alnap.org/resource/8127.aspx)).
- Hidalgo, S., LaGuardia, D., Trudi, G., Sole, R., Moussa, Z., van Dijk, J., Merckx, P. and Zimmer, L. (2015)** *Beyond humanitarian assistance? UNHCR and the response to Syrian refugees in Jordan and Lebanon: January 2013 – April 2014*. Brussels: Transtec. ([www.alnap.org/resource/20222.aspx](http://www.alnap.org/resource/20222.aspx)).
- Hidalgo, S. and Théodate, M. P. (2011)** *Inter-agency real-time evaluation of the humanitarian response to the earthquake in Haiti: 20 months after*. Geneva: IASC. ([www.alnap.org/resource/6330.aspx](http://www.alnap.org/resource/6330.aspx)).
- Hill, M. E. (2002)** 'Race of the interviewer and perception of skin color: Evidence from the multi-city study of urban inequality'. *American Sociological Review*, 67(1): 99–108. ([www.alnap.org/resource/8123.aspx](http://www.alnap.org/resource/8123.aspx)).
- Horton, K. and Roche, C. (eds) (2010)** *Ethical questions and international NGOs: An exchange between philosophers and NGOs*. Houten: Springer. ([www.alnap.org/resource/22955.aspx](http://www.alnap.org/resource/22955.aspx)).
- Hróbjartsson, A., Forfang, E., Haahr, M. T., Als-Nielsen, B. and Brorson, S. (2007)** 'Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding'. *International Journal of Epidemiology*, 36(3): 654–663. ([www.alnap.org/resource/22956.aspx](http://www.alnap.org/resource/22956.aspx)).

**Humanitarian Initiatives UK. (2001)** *Independent evaluation: The DEC response to the earthquake in Gujarat January - October 2001*. London: Humanitarian Initiatives UK. ([www.alnap.org/resource/3432.aspx](http://www.alnap.org/resource/3432.aspx)).

**Hund, L., Bedrick, E. J. and Pagano, M. (2015)** 'Choosing a cluster sampling design for lot quality assurance sampling surveys. *PLoS ONE*, 10(6). ([www.alnap.org/resource/22958.aspx](http://www.alnap.org/resource/22958.aspx)).

**Hunt, D. (2014)** *Photo Story: Risk Analysis in Tcharow, Goz Beida, Sila Region, Chad*. Concern Worldwide.

**Hyatt, S. and Auten, S. (2011)** *Evaluation of the Palestinian youth empowerment program (RUWWAD)*. Washington, DC: USAID. ([www.alnap.org/resource/22959.aspx](http://www.alnap.org/resource/22959.aspx)).

**IASC. (2010)** *Gender marker: How-to-code tip sheet*. Geneva: IASC. ([www.alnap.org/resource/8128.aspx](http://www.alnap.org/resource/8128.aspx)).

**IASC. (2010b)** *Guidance note for clusters to implement the IASC gender marker: Creating gender-responsive projects and tracking gender-related allocations in humanitarian appeals and funding mechanisms*. Geneva: IASC. ([www.alnap.org/resource/8129](http://www.alnap.org/resource/8129)).

**IASC. (2010c)** *IASC gender marker fact sheet*. Geneva: IASC. ([www.alnap.org/resource/22800.aspx](http://www.alnap.org/resource/22800.aspx)).

**IASC. (2010d)** *Inter-agency real-time evaluation (IA RTE) of the humanitarian response to [disaster XYZ In country XYZ]: Terms of reference*. Geneva: IASC. ([www.alnap.org/resource/22960.aspx](http://www.alnap.org/resource/22960.aspx)).

**IASC. (2011)** *IASC operational guidelines on protection of persons in situations of natural disasters*. Geneva: IASC. ([www.alnap.org/resource/12828.aspx](http://www.alnap.org/resource/12828.aspx)).

**IASC. (2013)** *The centrality of protection in humanitarian action*. Geneva: IASC. ([www.alnap.org/resource/19093.aspx](http://www.alnap.org/resource/19093.aspx)).

**ICAI. (2013)** *DFID's support for Palestine refugees through UNRWA*. London: ICAI. ([www.alnap.org/resource/10111.aspx](http://www.alnap.org/resource/10111.aspx)).

**ICRC. (1977)** *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of non-international armed conflicts (Protocol II)*. Geneva: ICRC. ([www.alnap.org/resource/8122.aspx](http://www.alnap.org/resource/8122.aspx)).

**IFAD. (2011)** *Evaluating the impact of participatory mapping activities*. Rome: IFAD. ([www.alnap.org/resource/22962.aspx](http://www.alnap.org/resource/22962.aspx)).

**IFAD. (2014)** *Calculating the sample size*. Rome: IFAD. ([www.alnap.org/resource/22961.aspx](http://www.alnap.org/resource/22961.aspx)).

**IFRC (n.d.)** *Types of disasters*. Geneva: IFRC. ([www.alnap.org/resource/22963.aspx](http://www.alnap.org/resource/22963.aspx)).

**IFRC. (2011)** *IFRC framework for evaluation*. Geneva: IFRC. ([www.alnap.org/resource/7975.aspx](http://www.alnap.org/resource/7975.aspx)).

**IFRC. (2010)** *World disasters report 2010: Focus on urban risk*. Geneva: IFRC. ([www.alnap.org/resource/9272.aspx](http://www.alnap.org/resource/9272.aspx)).

**IFRC. (2011)** *Project/programme monitoring and evaluation (M&E) guide*. Geneva: IFRC. ([www.alnap.org/resource/8542.aspx](http://www.alnap.org/resource/8542.aspx)).

**IFRC and ICRC. (1994)** *Code of conduct for the International Red Cross and Red Crescent movement and non-governmental organizations (NGOs) in disaster relief*. Geneva: IFRC/ICRC. ([www.alnap.org/resource/19291.aspx](http://www.alnap.org/resource/19291.aspx)).

**3ie. (2008a)** *3ie impact evaluation practice: a guide for grantees*. New Delhi: 3ie ([www.alnap.org/resource/8130.aspx](http://www.alnap.org/resource/8130.aspx)).

**3ie. (2008b)** *Principles for impact evaluation*. New Delhi: 3ie. ([www.alnap.org/resource/8131.aspx](http://www.alnap.org/resource/8131.aspx)).

**ILO. (2014a)** *Checklist 5 preparing the evaluation report*. Geneva: ILO. ([www.alnap.org/resource/22964.aspx](http://www.alnap.org/resource/22964.aspx)).

**ILO. (2014b)** *Checklist 6 rating the quality of evaluation*. Geneva: ILO. ([www.alnap.org/resource/22965.aspx](http://www.alnap.org/resource/22965.aspx)).

**Imbens, G. W. and Lemieux, T. (2007)** 'Regression discontinuity designs: A guide to practice'. *Journal of Econometrics*, 142(2): 615–635. ([www.alnap.org/resource/22966.aspx](http://www.alnap.org/resource/22966.aspx)).

**INEE. (2004)** *Minimum standards for education in emergencies, chronic crises and early reconstruction*. New York: INEE. ([www.alnap.org/resource/7860.aspx](http://www.alnap.org/resource/7860.aspx)).



**Innovation Network. (2005)** *Logic model workbook*. Washington, DC: Innovation Network. ([www.alnap.org/resource/22967.aspx](http://www.alnap.org/resource/22967.aspx)).

**Israel, G. D. (2009)** *Determining sample size*. Gainesville: IFAS University of Florida. ([www.alnap.org/resource/22968.aspx](http://www.alnap.org/resource/22968.aspx)).

**Jadad, A. R., Moore, A., Carroll, D., Jenkinson, C., Reynolds, J. M., Gavaghan, D. J. and McQuay, H. J. (1996)** 'Assessing the quality of reports of randomized clinical trials: is blinding necessary?' *Controlled Clinical Trials*, 17(1): 1–12. ([www.alnap.org/resource/22969.aspx](http://www.alnap.org/resource/22969.aspx)).

**Jalan, J. and Ravallion, M. (2001)** 'Does piped water reduce diarrhea for children in rural India?' *Journal of Econometrics*, 112(1): 153-173. ([www.alnap.org/resource/22970.aspx](http://www.alnap.org/resource/22970.aspx)).

**Jaspars, S. and Young, H. (1995)** *General food distribution in emergencies*, London: ODI. ([www.alnap.org/resource/2725.aspx](http://www.alnap.org/resource/2725.aspx)).

**Johnson, A. and Head, D. (2011)** *CRS Haiti real time evaluation of the 2010 earthquake response : findings, recommendations, and suggested follow up*. Baltimore, MD: CRS. ([www.alnap.org/resource/6074.aspx](http://www.alnap.org/resource/6074.aspx)).

**Johri, M., Ridde, V., Heinmüller, R. and Haddad, S. (2014)** *Estimation of maternal and child mortality one year after user-fee elimination: an impact evaluation and modelling study in Burkina Faso*. Geneva: WHO. ([www.alnap.org/resource/22971.aspx](http://www.alnap.org/resource/22971.aspx)).

**Jones, N., Jones, H., Steer, L. and Datta, A. (2009)** *Improving impact evaluation production and use*. London: ODI. ([www.alnap.org/resource/8225.aspx](http://www.alnap.org/resource/8225.aspx)).

**Journal of Peacebuilding & Development. (2013)** 'Evaluation in violently divided societies: Politics, ethics and methods'. *Journal of Peacebuilding & Development*, 8(2): 1–4. ([www.alnap.org/resource/23078.aspx](http://www.alnap.org/resource/23078.aspx)).

**Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1991)** 'The endowment effect, loss aversion, and status quo bias'. *The Journal of Economic Perspectives*, 5(1): 193–206. ([www.alnap.org/resource/22972.aspx](http://www.alnap.org/resource/22972.aspx)).

**Kaiser, R., Woodruff, B. A., Bilukha, O., Spiegel, P. B. and Salama, P. (2006)** 'Using design effects from previous cluster surveys to guide sample size calculation in emergency settings'. *Disasters*, 30(2): 199–211. ([www.alnap.org/resource/22973.aspx](http://www.alnap.org/resource/22973.aspx)).

**Kalantari, M., Yule, W., Dyregrov, A., Neshatdoost, H. and Ahmadi, S. J. (2012)** 'Efficacy of writing for recovery on traumatic grief symptoms of Afghani refugee bereaved adolescents: A randomized control trial. *OMEGA - Journal of Death and Dying*, 65(2): 139–150. ([www.alnap.org/resource/22974.aspx](http://www.alnap.org/resource/22974.aspx)).

**Kane, E. W. and Macaulay, L. J. (1993)** 'Interviewer gender and gender attitudes'. *Public Opinion Quarterly*, 57(1): 1–28. ([www.alnap.org/resource/8132.aspx](http://www.alnap.org/resource/8132.aspx)).

**Kaplan, A. M. and Haenlein, M. (2010)** 'Users of the world, unite! The challenges and opportunities of social media'. *Business Horizons*, 53(1): 59–68. ([www.alnap.org/resource/22975.aspx](http://www.alnap.org/resource/22975.aspx)).

**Kennedy, M. (2006)** *Interview Probes*. ([www.alnap.org/resource/22976.aspx](http://www.alnap.org/resource/22976.aspx)).

**Kevlihan, R. (2010)** Productivity and cash-for-work in Niger: GOAL's experience. *Humanitarian Exchange* 48: 29-30. London: HPN/ODI. ([www.alnap.org/resource/8135.aspx](http://www.alnap.org/resource/8135.aspx)).

**Khandker, S. R., Koolwal, G. B. and Samad, H. A. (2010)** *Handbook on impact evaluation: quantitative methods and practices*. Washington, DC: World Bank. ([www.alnap.org/resource/22977.aspx](http://www.alnap.org/resource/22977.aspx)).

**Khoo, S.-E. (2010)** 'Health and humanitarian migrants' economic participation. *Journal of Immigrant and Minority Health*, 12(3): 327–339. ([www.alnap.org/resource/22978.aspx](http://www.alnap.org/resource/22978.aspx)).

**King, J., McKegg, K., Oakden, J. and Wehipeihana, N. (2013)** 'Rubrics: A method for surfacing values and improving the credibility of evaluation'. *Journal of multidisciplinary evaluation*, 9(21): 11–20. ([www.alnap.org/resource/22424.aspx](http://www.alnap.org/resource/22424.aspx)).

**Kirkby, J., Saeed, A. and Zogopoulos, A. (2006)** *Independent evaluation of CARE international's earthquake response in Northern Pakistan*. Atlanta, GA: CARE International. ([www.alnap.org/resource/3489.aspx](http://www.alnap.org/resource/3489.aspx)).

**Knox Clarke, P. and Darcy, J. (2014)** *Insufficient evidence? The quality and use of evidence in humanitarian action*. ALNAP Study. London: ALNAP/ODI. ([www.alnap.org/resource/10441.aspx](http://www.alnap.org/resource/10441.aspx)).

**Koenig, T. (2004)** *CAQDAS - A primer*. ([www.alnap.org/resource/22979.aspx](http://www.alnap.org/resource/22979.aspx)).

**Kohrt, B. A., Burkey, M., Stuart, E., A. and Koirala, S. (2015)** 'Alternative approaches for studying humanitarian interventions: propensity score methods to evaluate reintegration packages impact on depression, PTSD, and function impairment among child soldiers in Nepal'. *Global Mental Health*. (www.alnap.org/resource/22980.aspx).

**Kosko, J., Klassen, T. P., Bishop, T. and Hartling, L. (2006)** 'Evidence-based medicine and the anecdote: Uneasy bedfellows or ideal couple?' *Paediatrics & Child Health*, 11(10): 665–668. (www.alnap.org/resource/22981.aspx).

**Krippendorff, K. (2004)** *Content analysis: an introduction to its methodology*. 2nd ed. Thousand Oaks, CA: SAGE. (www.alnap.org/resource/8223.aspx).

**Krueger, R. A. and Casey, M. A. (2009)** *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: Sage. (www.alnap.org/resource/8136.aspx).

**Krueger, S. and Sagmeister, E. (2014)** *Real-time evaluation of humanitarian assistance revisited: Lessons learned and the way forward*. Berlin: GPPi. (www.alnap.org/resource/12886.aspx).

**Kumar, S. (2002)** *Methods for community participation: a complete guide for practitioners*. Rugby: ITDG Publishing. (www.alnap.org/resource/8137.aspx).

**Kusters, C., van Vugt, S., Wigboldus, S., Williams, B. and Woodhill, J. (2011)** *Making evaluations matter: A practical guide for evaluators*. (www.alnap.org/resource/22425.aspx).

**Ky Luu, J. D., Nkwake, A. M., Morris, M. F., Hansch, S., Gago, M., Corzine, E. and Marcello, J. (2014)** *A decade of learning: Lessons from an evaluation of the emergency capacity building (ECB) project*. New Orleans, LA: DRG. (www.alnap.org/resource/12981.aspx).

**LaGuardia, D. and Van den Toorn, W. (2011)** *Evaluation of UNRWA's organizational development (OD)*. Amman: UNRWA. (www.alnap.org/resource/22986.aspx).

**Lanphear, B. P., Vorhees, C. V. and Bellinger, D. C. (2005)** 'Protecting children from environmental toxins'. *PLoS Medicine*, 2(3): 203-208. (www.alnap.org/resource/22987.aspx).

- Larcom, S., Rauch, F. and Willems, T. (2015)** *The benefits of forced experimentation: Striking evidence from the London Underground network*, Oxford: University of Oxford. ([www.alnap.org/resource/22988.aspx](http://www.alnap.org/resource/22988.aspx)).
- Lautze, S. and Raven-Roberts, A. (2003)** *Food security in complex emergencies: building policy frameworks to address longer-term programming challenges*. Rome: FAO. ([www.alnap.org/resource/7785.aspx](http://www.alnap.org/resource/7785.aspx)).
- Lawday, A. (2014)** *Evaluation of HelpAge International's programme "assistance to specific vulnerable groups affected by the Syrian crisis" (2013-2014)*. London: Andrew Lawday Evaluations. ([www.alnap.org/resource/12857.aspx](http://www.alnap.org/resource/12857.aspx)).
- Lawday, A. (2015)** *Inter-agency humanitarian evaluation (IAHE) of the response to the Central African Republic's crisis 2013-2015*. Washington: Konterra Group. ([www.alnap.org/resource/22989.aspx](http://www.alnap.org/resource/22989.aspx)).
- Laybourn, C. and Obrecht, A. (2014)** *DEC accountability self-assessment validation 2013/14*. London: One World Trust. ([www.alnap.org/resource/22990.aspx](http://www.alnap.org/resource/22990.aspx)).
- Leader, N. (2000)** *The politics of principle: the principles of humanitarian action in practice*. London: ODI/HPG. ([www.alnap.org/resource/8139.aspx](http://www.alnap.org/resource/8139.aspx)).
- Lee, R. M. (2000)** *Unobtrusive methods in social research*. Buckingham: Open University Press. ([www.alnap.org/resource/22991.aspx](http://www.alnap.org/resource/22991.aspx)).
- Lemeshow, S. and Taber, S. (1991)** 'Lot quality assurance sampling: single-and double-sampling plans'. *World Health Stat Q*, 44(3): 115–132. ([www.alnap.org/resource/22992.aspx](http://www.alnap.org/resource/22992.aspx)).
- Leturque, H., Beaujeu, R., Majeed, Y. and Saleem, S. (2012)** *Evaluation of the CBHA early recovery programme in Pakistan*. London: CBHA. ([www.alnap.org/resource/6370.aspx](http://www.alnap.org/resource/6370.aspx)).
- Lieberson, S. (1991)** 'Small N's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases'. *Social Forces*, 70(2): 307–320. ([www.alnap.org/resource/22994.aspx](http://www.alnap.org/resource/22994.aspx)).
- Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H. van der Meulen, J. H. P. and Bossuyt, P. M. M. (1999)** 'Empirical evidence of design-related bias in studies of diagnostic tests'. *JAMA*, 282(11): 1061–1066. ([www.alnap.org/resource/22995.aspx](http://www.alnap.org/resource/22995.aspx)).

- Lindgren, D., Matondang, M. and Putri, D. (2005)** *Tsunami relief study: Effectiveness of the tsunami relief effort*. London: TEC. ([www.alnap.org/resource/22997.aspx](http://www.alnap.org/resource/22997.aspx)).
- Lister, S., Anderson, S., Mokbel Genequand, M., Gordon, A., Sandford, J. and Turner, S. (2011)** *WFP's school feeding policy: a policy evaluation. Vol. 1 full report*. Rome: WFP. ([www.alnap.org/resource/22998.aspx](http://www.alnap.org/resource/22998.aspx)).
- Longo, M., Canetti, D. and Hite-Rubin, N. (2014)** 'A checkpoint effect? Evidence from a natural experiment on travel restrictions in the West Bank'. *American Journal of Political Science*, 58(4): 1006-1023. ([www.alnap.org/resource/23000.aspx](http://www.alnap.org/resource/23000.aspx)).
- Lysy, C. (2014)** *More data, shorter reports: three ways to respond*. Fresh Spectrum ([www.alnap.org/resource/23001.aspx](http://www.alnap.org/resource/23001.aspx)).
- MacDonald, B. and Gedeon, H. (2012)** *Banking with mobile phone: A report on a T-cash pilot project*, Baltimore, MD: CRS. ([www.alnap.org/resource/9882.aspx](http://www.alnap.org/resource/9882.aspx)).
- Mackay, K., Clark, M., Sartorius, R., Bamberger, M. (2004)** *Monitoring & evaluation: Some tools, methods & approaches*. Washington, DC: World Bank. ([www.alnap.org/resource/23003.aspx](http://www.alnap.org/resource/23003.aspx)).
- Mahoney, J. (2000)** 'Strategies of causal inference in small-N analysis'. *Sociological Methods & Research*, 28(4): 387–424. ([www.alnap.org/resource/23004.aspx](http://www.alnap.org/resource/23004.aspx)).
- Mahoney, J. and Goertz, G. (2006)** 'A tale of two cultures: Contrasting quantitative and qualitative research'. *Political Analysis*, 14(3): 227–249. ([www.alnap.org/resource/22637.aspx](http://www.alnap.org/resource/22637.aspx)).
- Majewski, B., Boulet-Desbureau, P., Slezak, M., De Meulder, F. and Wilson, K. (2012)** *Joint evaluation of the global logistics cluster*. ([www.alnap.org/resource/7904.aspx](http://www.alnap.org/resource/7904.aspx)).
- Malki, E. (2008)** *A benchmarking model for measuring the efficiency of a humanitarian aid program: a case study of an international NGO*. Munich: MPRA. ([www.alnap.org/resource/23006.aspx](http://www.alnap.org/resource/23006.aspx)).
- Manesh, A. O., Sheldon, T., A., Pickett, K. E., and Carr-Hill, R. (2008)** 'Accuracy of child morbidity data in demographic and health surveys'. *International Journal of Epidemiology*, 37(1). ([www.alnap.org/resource/8140.aspx](http://www.alnap.org/resource/8140.aspx)).

- Manning, R. (2009)** *The quality of DFID's evaluations and evaluation management systems: How do they compare with other agencies?* London: IACDI. ([www.alnap.org/resource/23007.aspx](http://www.alnap.org/resource/23007.aspx)).
- UN OIOS. (2014)** *Inspection and evaluation manual*. New York: UN OIOS. ([www.alnap.org/resource/19243.aspx](http://www.alnap.org/resource/19243.aspx)).
- Marshall, M. N. (1996)** 'Sampling for qualitative research'. *Family Practice*, 13(6): 522–526. ([www.alnap.org/resource/23008.aspx](http://www.alnap.org/resource/23008.aspx)).
- Martin, E., Petty, C. and Acidri, J. (2009)** *Livelihoods in crisis: a longitudinal study in Pader, Uganda*. London: HPG/ODI. ([www.alnap.org/resource/23010.aspx](http://www.alnap.org/resource/23010.aspx)).
- Martinez, A. (2009)** *A participatory programme review of the international federation of the Red Cross and Red Crescent societies response to the Peru earthquake of 15th August 2007*. Geneva: IFRC. ([www.alnap.org/resource/8141.aspx](http://www.alnap.org/resource/8141.aspx)).
- Martinez, D. E. (2011)** *The logical framework approach in non-governmental organizations*. Edmonton: University of Alberta. ([www.alnap.org/resource/23011.aspx](http://www.alnap.org/resource/23011.aspx)).
- Mays, N. and Pope, C. (2000)** 'Assessing quality in qualitative research'. *BMJ*. ([www.alnap.org/resource/8142.aspx](http://www.alnap.org/resource/8142.aspx)).
- Maystadt, J. F. (2011)** *Poverty reduction in a refugee-hosting economy: A natural experiment*. Washington, DC: IFPRI. ([www.alnap.org/resource/8143.aspx](http://www.alnap.org/resource/8143.aspx)).
- Mazurana, D., Benelli, P., Gupta, H. and Walker, P. (2011)** *Sex and age matter: Improving humanitarian response in emergencies*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8144.aspx](http://www.alnap.org/resource/8144.aspx)).
- McDonald, B. and Rogers, P. (2014)** *Interviewing*. Florence: UNICEF. ([www.alnap.org/resource/23012.aspx](http://www.alnap.org/resource/23012.aspx)).
- McGearty, S., O'Hagan, P. and Montinard, M. (2012)** *An independent final evaluation of the action of churches together alliance Haiti appeal HTI-101 (Jan 2010 – Dec 2011)*. London: Channel Research. ([www.alnap.org/resource/6339.aspx](http://www.alnap.org/resource/6339.aspx)).
- McRAM Team. (2009)** *Multi-cluster rapid assessment mechanism (McRAM) in Pakistan*. Islamabad: McRAM. ([www.alnap.org/resource/8145.aspx](http://www.alnap.org/resource/8145.aspx)).

**Hanif, A. (2008)** *HAP initiatives in Concern Worldwide: A case study of Bangladesh (2008)*. London: Concern Worldwide. ([www.alnap.org/resource/23013.aspx](http://www.alnap.org/resource/23013.aspx)).

**Mendoza, U. and Thomas, V. (2009)** *UNHCR's AGDM evaluation: A participatory evaluation of AGDM results in four Colombian communities*. Geneva: UNHCR. ([www.alnap.org/node/24503.aspx](http://www.alnap.org/node/24503.aspx)).

**Meng, X. and Qian, N. (2009)** *The long term consequences of famine on survivors: Evidence from a unique natural experiment using China's great famine*. Cambridge, MA: NBER. ([www.alnap.org/resource/23015.aspx](http://www.alnap.org/resource/23015.aspx)).

**Meyer, A.-M. (2007)** *The causes of malnutrition in children under 3 in the Somali Region of Ethiopia related to household caring practices*. London: Save the Children. ([www.alnap.org/resource/3591.aspx](http://www.alnap.org/resource/3591.aspx)).

**Mezuk, B., Larkin, G. L., Prescott, M. R., Tracy, M., Vlahov, D., Tardiff, K. and Galea, S. (2009)** 'The influence of a major disaster on suicide risk in the population'. *Journal of Traumatic Stress*, 22(6): 481-488. ([www.alnap.org/resource/23017.aspx](http://www.alnap.org/resource/23017.aspx)).

**Miles, M. B., Huberman, A. M. and Saldaña, J. (2013)** *Qualitative data analysis: A methods sourcebook*. 3rd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/23018.aspx](http://www.alnap.org/resource/23018.aspx)).

**Miller, J. and Stolz, K. (2008)** *Moving beyond rhetoric: Consultation and participation with populations displaced by conflict or natural disasters*. Washington, DC: The Brookings Institution. ([www.alnap.org/resource/8751.aspx](http://www.alnap.org/resource/8751.aspx)).

**Milojevic, A., Armstrong, B., Hashizume, M., McAllister, K., Faruque, A., Yunus, M., Kim Streatfield, P. Moji, K. and Wilkinson, P. (2012)** 'Health effects of flooding in rural Bangladesh'. *Epidemiology*, 23(1): 107–115. ([www.alnap.org/resource/23019.aspx](http://www.alnap.org/resource/23019.aspx)).

**Milojevic, A., Armstrong, B., Kovats, S., Butler, B., Hayes, E., Leonardi, G., Murray, V. and Wilkinson, P. (2011)** 'Long-term effects of flooding on mortality in England and Wales, 1994-2005: controlled interrupted time-series analysis'. *Environmental Health*. ([www.alnap.org/resource/23020.aspx](http://www.alnap.org/resource/23020.aspx)).

**Miranda, R., (2009)** *Eva the Evaluator*. Learningham Press. ([www.alnap.org/resource/23028.aspx](http://www.alnap.org/resource/23028.aspx)).

- Mirano, S. (2012)** *Learning from the Urban Transitional Shelter Response in Haiti*. Baltimore, MD: CRS. ([www.alnap.org/resource/6539.aspx](http://www.alnap.org/resource/6539.aspx)).
- Molund, S. and Schill, G. (2007)** *Looking back, moving forward*. 2nd ed. Stockholm: SIDA. ([www.alnap.org/resource/7902.aspx](http://www.alnap.org/resource/7902.aspx)).
- Molyneux, S., Mulupi, S., Mbaabu, L. and Marsh, V. (2012)** 'Benefits and payments for research participants: Experiences and views from a research centre on the Kenyan coast'. *BMC Medical Ethics*, 13(1). ([www.alnap.org/resource/8147.aspx](http://www.alnap.org/resource/8147.aspx)).
- Montbiot, E. (2006)** *World Vision: Lessons by Sectors*. Monrovia, CA: World Vision. ([www.alnap.org/resource/23029.aspx](http://www.alnap.org/resource/23029.aspx)).
- Morán, J. L. Á. (2012)** *External evaluation: Emergency nutrition programme in Sindh Province, Pakistan*. London: Action Against Hunger. ([www.alnap.org/resource/20934.aspx](http://www.alnap.org/resource/20934.aspx)).
- Morel, D. and Hagens, C. (2012)** *Monitoring, evaluation, accountability and learning in emergencies - a resource pack for simple and strong MEAL*. Baltimore, MD: CRS. ([www.alnap.org/resource/9200.aspx](http://www.alnap.org/resource/9200.aspx)).
- Morgan, D. L. (1997)** *Focus groups as qualitative research*. 2nd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/23030.aspx](http://www.alnap.org/resource/23030.aspx)).
- Morinière, L. (2011a)** *Evaluation of the ERRF component of the Haiti emergency response (ERF) fund*. New York: OCHA. ([www.alnap.org/resource/6054.aspx](http://www.alnap.org/resource/6054.aspx)).
- Morinière, L. (2011b)** *External evaluation of the Haiti emergency relief and response fund (ERRF), 2008-2011*. New York: OCHA. ([www.alnap.org/resource/8149.aspx](http://www.alnap.org/resource/8149.aspx)).
- Morra-Imas, L. G. and Rist, R. C. (2009)** *The road to results: Designing and conducting effective development evaluations*. Washington, DC: World Bank. ([www.alnap.org/resource/8470.aspx](http://www.alnap.org/resource/8470.aspx)).
- Morris, M. F. and Shaughnessy, D. E. (2007)** *Final evaluation report: Emergency capacity building project*. Atlanta, GA: CARE International. ([www.alnap.org/resource/10636.aspx](http://www.alnap.org/resource/10636.aspx)).
- Morris, T., Steen, N. and Canteli, C. (2013)** *Synthesis of mixed method impact evaluations of the contribution of food assistance to durable solutions in protracted refugee situations*. Geneva: UNHCR. ([www.alnap.org/resource/19902.aspx](http://www.alnap.org/resource/19902.aspx)).



**Moscoe, E., Bor, J. and Bärnighausen, T. (2015)** 'Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice'. *Journal of Clinical Epidemiology*, 68(2): 132–143. (www.alnap.org/resource/23031.aspx).

**Mowjee, T., Fleming, D. and Toft, E. (2015)** *Evaluation of the strategy for Danish humanitarian action 2010-2015*. Copenhagen: DANIDA. (www.alnap.org/resource/20738.aspx).

**MSF, 2010.** *Summary of findings of emergency response evaluations*. Geneva: MSF. (www.alnap.org/resource/23032.aspx).

**MSF. (2012)** *A handbook for initiating, managing and conducting evaluations in MSF*. Geneva: MSF. (www.alnap.org/resource/10004.aspx).

**Must, A., Phillips, S., Naumova, E., Blum, M., Harris, S., Dawson-Hughes, B. and Rand, W. (2002)** 'Recall of early menstrual history and menarcheal body size: After 30 years, how well do women remember?' *American Journal of Epidemiology*, 155(7): 672–679. (www.alnap.org/resource/8150.aspx).

**La Pelle, N. R. (2004)** 'Simplifying qualitative data analysis using general purpose software tools'. *Field Methods*, 16(1): 85–108. (www.alnap.org/resource/23033.aspx).

**Naidu, V. J. (2010)** *A thematic evaluation of livelihood and community based disaster preparedness projects*. Atlanta, GA: CARE International. (www.alnap.org/resource/10252.aspx).

**NAO. (2000)** *A practical guide to sampling*. London: NAO. (www.alnap.org/resource/10228.aspx).

**NORAD. (1998)** *Results management in Norwegian development cooperation : A practical guide*. Oslo: NORAD. (www.alnap.org/resource/8052.aspx).

**NORAD. (1999)** *Logical framework approach: handbook for objectives-oriented planning*. Oslo: NORAD. (www.alnap.org/resource/23034.aspx).

**Norman, B. (2012)** *Monitoring and accountability practices for remotely managed projects implemented in volatile operating environments*. Teddington: Tearfund. (www.alnap.org/resource/7956.aspx).

**NRC. (2013)** *Evaluation of five humanitarian programmes of the Norwegian Refugee Council (NRC) and of the standby roster NORCAP. Case country report – Somalia.* Oslo: NRC. ([www.alnap.org/resource/12227.aspx](http://www.alnap.org/resource/12227.aspx)).

**O'Hagan, P., Love, K. and Rouse, A. (2010)** *An independent joint evaluation of the Haiti earthquake humanitarian response.* Atlanta, GA: CARE International. ([www.alnap.org/resource/5969.aspx](http://www.alnap.org/resource/5969.aspx)).

**O'Neil, G. (2012)** *7 new ways to present evaluation findings.* 3-5 October, Helsinki. ([www.alnap.org/resource/23035.aspx](http://www.alnap.org/resource/23035.aspx)).

**Oakden, J. (2013a)** *Evaluation rubrics: how to ensure transparent and clear assessment that respects diverse lines of evidence.* Melbourne: Better Evaluation. ([www.alnap.org/resource/23036.aspx](http://www.alnap.org/resource/23036.aspx)).

**Oakden, J., 2013b.** *Guest blog: Why rubrics are useful in evaluations.* Melbourne: Better Evaluation. ([www.alnap.org/resource/23037.aspx](http://www.alnap.org/resource/23037.aspx)).

**Obrecht, A., Laybourne, C., Hammer, M. and Ray, S. (2012)** *The 2011/12 DEC accountability framework assessment: findings from the external evaluation and peer review process.* London: One World Trust. ([www.alnap.org/resource/23038.aspx](http://www.alnap.org/resource/23038.aspx)).

**OCHA. (2004)** *Guiding principles on internal displacement.* New York: OCHA. ([www.alnap.org/resource/8054.aspx](http://www.alnap.org/resource/8054.aspx)).

**OCHA. (2009)** 'OCHA on message : Humanitarian access case study'. *OCHA annual report 2008.* New York: OCHA. ([www.alnap.org/resource/11252.aspx](http://www.alnap.org/resource/11252.aspx)).

**OCHA. (2011)** *Inter-agency RTE for the 2011 Pakistan floods.* New York: OCHA. ([www.alnap.org/resource/23040.aspx](http://www.alnap.org/resource/23040.aspx)).

**OCHA. (2012)** *OCHA on message: humanitarian principles.* New York: OCHA. ([www.alnap.org/resource/19270.aspx](http://www.alnap.org/resource/19270.aspx)).

**OCHA. (2013)** *United Nations Disaster Assessment and Coordination UNDAC field handbook.* New York: OCHA. ([www.alnap.org/resource/23041.aspx](http://www.alnap.org/resource/23041.aspx)).

**ODI. (2013)** *Research and policy in development (RAPID).* London: ODI. ([www.alnap.org/resource/23048.aspx](http://www.alnap.org/resource/23048.aspx)).

- OECD. (2010)** *Managing Joint Evaluations*. Paris: OECD.  
([www.alnap.org/resource/23039.aspx](http://www.alnap.org/resource/23039.aspx)).
- OECD/DAC. (1991)** *Principles for evaluation of development assistance*. Paris: OECD/DAC. ([www.alnap.org/resource/20830.aspx](http://www.alnap.org/resource/20830.aspx)).
- OECD/DAC. (1999)** *Guidance for evaluating humanitarian assistance in complex emergencies*. Paris: OECD/DAC. ([www.alnap.org/resource/8221.aspx](http://www.alnap.org/resource/8221.aspx)).
- OECD/DAC. (2010)** *Glossary of key terms in evaluation and results based management*. Paris: OECD/DAC. ([www.alnap.org/resource/8489.aspx](http://www.alnap.org/resource/8489.aspx)).
- OECD/DAC. (2012)** *Evaluating peacebuilding activities in settings of conflict and fragility: Improving learning for results*. Paris: OECD/DAC. ([www.alnap.org/resource/8057.aspx](http://www.alnap.org/resource/8057.aspx)).
- OIOS. (2014)** *Inspection and evaluation manual*. New York: OIOS.  
([www.alnap.org/resource/19243.aspx](http://www.alnap.org/resource/19243.aspx)).
- Ojha, G. P. (2010)** *Appreciative inquiry approach to evaluation practices in South Asia*. Iasi: Lumen Publishing House. ([www.alnap.org/resource/23042.aspx](http://www.alnap.org/resource/23042.aspx)).
- Oliver, M. L. (2007)** *CARE's humanitarian operations: Review of CARE's use of evaluations and after action reviews in decision-making*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8058.aspx](http://www.alnap.org/resource/8058.aspx)).
- Oliver, M. L. (2009)** 'Metaevaluation as a means of examining evaluation influence'. *Journal of MultiDisciplinary Evaluation*, 6(11): 32–37. ([www.alnap.org/resource/8285.aspx](http://www.alnap.org/resource/8285.aspx)).
- Ong, J. C., Flores, J. M. and Combinido, P. (2015)** *Obligated to be grateful*. Woking: Plan International. ([www.alnap.org/resource/20633.aspx](http://www.alnap.org/resource/20633.aspx)).
- Onwuegbuzie, A. J. and Collins, K. M. T. (2007)** 'A typology of mixed methods sampling designs in social science research'. *The Qualitative Report*, 12(2): 281–316. ([www.alnap.org/resource/23044.aspx](http://www.alnap.org/resource/23044.aspx)).
- Onwuegbuzie, A. J. and Leech, N. (2007)** 'A call for qualitative power analyses'. *Quality & Quantity*, 41(1): 105–121. ([www.alnap.org/resource/23045.aspx](http://www.alnap.org/resource/23045.aspx)).

**Ott, E., Krystalli, R. C., Stites, E., Timmins, N., Walden, V., Gillingham, E., Bushby, K. (2015)** *Humanitarian evidence synthesis and communication programme: Abridged inception report*. Oxford: Oxfam and Feinstein International Center. ([www.alnap.org/resource/23046.aspx](http://www.alnap.org/resource/23046.aspx)).

**Otto, R. (2015)** *Story in 5: Ralf Otto, Momologue and Ebaix*. [YouTube] London: ALNAP. ([www.alnap.org/resource/23047.aspx](http://www.alnap.org/resource/23047.aspx)).

**Otto, R., Conrad, C., Brusset, E. and Stone, L. (2006)** *Organisational learning review of Caritas Internationalis' response to the Tsunami emergency*. London: Channel Research. ([www.alnap.org/resource/8059.aspx](http://www.alnap.org/resource/8059.aspx)).

**Oxfam. (2011)** *How are effectiveness reviews carried out?* Oxford: Oxfam. ([www.alnap.org/resource/8473.aspx](http://www.alnap.org/resource/8473.aspx)).

**Oxfam. (2012)** *Project effectiveness reviews: summary of 2011/12 findings and lessons learned*. Oxford: Oxfam. ([www.alnap.org/resource/8472.aspx](http://www.alnap.org/resource/8472.aspx)).

**Page, L., Savage, D. and Torgler, B. (2012)** *Variation in risk seeking behavior in a natural experiment on large losses induced by a natural disaster*. Zürich: CREMA. ([www.alnap.org/resource/23050.aspx](http://www.alnap.org/resource/23050.aspx)).

**Pandya, M., Pathak, V. and Oza, S. (2012)** *Future of M&E in humanitarian sector: Possible list for discussion*. Gujarat: AIDMI. ([www.alnap.org/resource/22894.aspx](http://www.alnap.org/resource/22894.aspx)).

**Pantera, G. (2012)** *AGIRE humanitarian response to the East Africa drought*. Rome: AGIRE. ([www.alnap.org/resource/6316.aspx](http://www.alnap.org/resource/6316.aspx)).

**Patrick, J. (2011)** *Haiti earthquake response. Emerging evaluation lessons*. London: ALNAP. ([www.alnap.org/resource/6125.aspx](http://www.alnap.org/resource/6125.aspx)).

**Patton, M. Q. (1997)** *Utilisation-focused evaluation: the new century text*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/11131.aspx](http://www.alnap.org/resource/11131.aspx)).

**Patton, M. Q. (2008)** *Utilization-focused evaluation*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/10060.aspx](http://www.alnap.org/resource/10060.aspx)).

**Patton, M. Q. and Cochran, M. (2002)** *A guide to using qualitative research methodology*. Geneva: MSF. ([www.alnap.org/resource/13024.aspx](http://www.alnap.org/resource/13024.aspx)).

- Pawson, R. and Tilley, N. (2004)** *Realist evaluation*. Mount Torrens: Community Matters. ([www.alnap.org/resource/8195.aspx](http://www.alnap.org/resource/8195.aspx)).
- Peersman, G. (2014a)** *Evaluative criteria*. Florence: UNICEF. ([www.alnap.org/resource/23051.aspx](http://www.alnap.org/resource/23051.aspx)).
- Peersman, G. (2014b)** *Overview: Data collection and analysis methods in impact evaluation*. Florence: UNICEF. ([www.alnap.org/resource/22654.aspx](http://www.alnap.org/resource/22654.aspx)).
- Penfold, R. B. and Zhang, F. (2013)** 'Use of interrupted time series analysis in evaluating health care quality improvements'. *Academic Pediatrics*, 13(6): S38–S44. ([www.alnap.org/resource/23052.aspx](http://www.alnap.org/resource/23052.aspx)).
- People in Aid. (2003)** *Code of good practice in the management and support of aid personnel*. London: People in Aid. ([www.alnap.org/resource/8061.aspx](http://www.alnap.org/resource/8061.aspx)).
- Perrin, B. (2012)** *Linking Monitoring and Evaluation to Impact Evaluation*. Washington, DC: InterAction. ([www.alnap.org/resource/19267.aspx](http://www.alnap.org/resource/19267.aspx)).
- Petticrew, M., Cummins, S., Ferrell, C., Findlay, A., Higgins, C. Hoy, C., Kearns, A. and Sparks, L. (2005)** 'Natural experiments: an underused tool for public health?' *Public Health*, 119(9): 751–757. ([www.alnap.org/resource/23053.aspx](http://www.alnap.org/resource/23053.aspx)).
- Pfister, R. (2011)** 'Wardrobe malfunctions and the measurement of internet behaviour'. *Psychology*, 2(3): 266–268. ([www.alnap.org/resource/23054.aspx](http://www.alnap.org/resource/23054.aspx)).
- Picciotto, R. (2012)** 'Experimentalism and development evaluation: Will the bubble burst?' *Evaluation*, 18(2): 213–229. ([www.alnap.org/resource/8062.aspx](http://www.alnap.org/resource/8062.aspx)).
- Rawlins, R., Pimkina, S., Barrett, C. B., Pedersen, S. and Wydick, B. (2013)** 'Got milk? The impact of Heifer International's livestock donation programs in Rwanda'. *Food Policy*, 44(2): 202–213. ([www.alnap.org/resource/23055.aspx](http://www.alnap.org/resource/23055.aspx)).
- Polastro, R. (2014)** *Evaluating humanitarian action in real time: Recent practices, challenges, and innovations*. Sheffield: IOD Parc. ([www.alnap.org/resource/12928.aspx](http://www.alnap.org/resource/12928.aspx)).
- Polastro, R., Khalif, M. A., van Ebyen, M., Posada, S., Salah, A., Steen, N. and Toft, E. (2011)** *IASC evaluation of the humanitarian response in South Central Somalia 2005-2010*. Madrid: DARA. ([www.alnap.org/resource/9301.aspx](http://www.alnap.org/resource/9301.aspx)).

**Polastro, R., Nagrah, A., Steen, N. and Zafar, F. (2011)** *Inter-agency real time evaluation of the humanitarian response to Pakistan's 2010 Flood Crisis*. Madrid: DARA. ([www.alnap.org/resource/6087.aspx](http://www.alnap.org/resource/6087.aspx)).

**Poulos, C., Pattanayak, S. K. and Jones, K. (2006)** *A guide to water and sanitation sector impact evaluations*. Washington, DC: World Bank. ([www.alnap.org/resource/8063.aspx](http://www.alnap.org/resource/8063.aspx)).

**Poulsen, L., Stacy, R., Bell, L. and Kumar Range, S. (2009)** *Joint thematic evaluation of FAO and WFP support to information systems for food security*. Rome: WFP. ([www.alnap.org/resource/5845.aspx](http://www.alnap.org/resource/5845.aspx)).

**Preskill, H. S. and Catsambas, T. T. (2006)** *Reframing evaluation through appreciative inquiry*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8224.aspx](http://www.alnap.org/resource/8224.aspx)).

**Proudlock, K., Ramalingam, B. and Sandison, P. (2009)** 'Improving humanitarian impact assessment: bridging theory and practice', in *ALNAP's 8th review of humanitarian action: Performance, Impact and Innovation*. ALNAP Review. London: ALNAP. ([www.alnap.org/resource/5663.aspx](http://www.alnap.org/resource/5663.aspx)).

**Puoane, T., Alexander, L. and Hutton, B. (2011)** *The revolving door: child malnutrition in Mount Frere, Eastern Cape Province of South Africa*. Cape Town: University of the Western Cape. ([www.alnap.org/resource/23056.aspx](http://www.alnap.org/resource/23056.aspx)).

**Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y. and Brück, T. (2014)** *What methods may be used in impact evaluations of humanitarian assistance?* New Delhi: 3ie. ([www.alnap.org/resource/19288.aspx](http://www.alnap.org/resource/19288.aspx)).

**Qasim, Q., 3ie and Stevens, K. (2014)** *Regression Discontinuity*. Melbourne: Better Evaluation. ([www.alnap.org/resource/23057.aspx](http://www.alnap.org/resource/23057.aspx)).

**Rabbani, M., Prakash, V. A. and Sulaiman, M. (2006)** *Impact assessment of CFPR/TUP: a descriptive analysis based on 2002-2005 panel data*. Dhaka: BRAC. ([www.alnap.org/resource/23058.aspx](http://www.alnap.org/resource/23058.aspx)).

**Rademaker, L. L., Grace, E. J. and Curda, S. K. (2012)** 'Using computer-assisted qualitative data analysis software (CAQDAS) to re-examine traditionally analyzed data: Expanding our understanding of the data and of ourselves as scholars'. *The Qualitative Report*, 17(22): 1–11. ([www.alnap.org/resource/23059.aspx](http://www.alnap.org/resource/23059.aspx)).

**Ramalingam, B. (2011)** *Learning how to learn: eight lessons for impact evaluations that make a difference*. London: ODI. ([www.alnap.org/resource/8458.aspx](http://www.alnap.org/resource/8458.aspx)).

**Ramalingam, B. and Mitchell, J. (2014)** *Responding to changing needs?* London: ALNAP. ([www.alnap.org/resource/19246.aspx](http://www.alnap.org/resource/19246.aspx)).

**Raphael, K. (1987)** 'Recall bias: A proposal for assessment and control'. *International Journal of Epidemiology*, 16(2): 167–170. ([www.alnap.org/resource/23060.aspx](http://www.alnap.org/resource/23060.aspx)).

**Raworth, K. (2016)** *Reviewing the existing literature*. Oxford: Oxfam. ([www.alnap.org/resource/23049.aspx](http://www.alnap.org/resource/23049.aspx)).

**Reddit Inc. (2013)** *Reddit: the front page of the internet*. ([www.alnap.org/resource/23061.aspx](http://www.alnap.org/resource/23061.aspx)).

**Reid, E. and Novak, P. (1975)** 'Personal space: An unobtrusive measures study'. *Bulletin of the Psychonomic Society*, 5(3): 265–266. ([www.alnap.org/resource/23062.aspx](http://www.alnap.org/resource/23062.aspx)).

**Reid, S., Stibbe, D. and Lowery, C. (2010)** *Review of the global education cluster co-leadership arrangement*. Oxford: The Partnering Initiative. ([www.alnap.org/resource/6354.aspx](http://www.alnap.org/resource/6354.aspx)).

**Remler, D. K. and Van Ryzin, G. G. (2014)** *Research methods in practice: Strategies for description and causation*. 2nd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/23063.aspx](http://www.alnap.org/resource/23063.aspx)).

**Ridde, V., Haddad, S. and Heinmüller, R. (2013)** 'Improving equity by removing healthcare fees for children in Burkina Faso'. *Journal of Epidemiology and Community Health*. 2013(0): 1-7. ([www.alnap.org/resource/23064.aspx](http://www.alnap.org/resource/23064.aspx)).

**Riddell, R. C. (2009)** *The quality of DFID's evaluation reports and assurance systems: A report for IACDI based on the quality review undertaken by consultants Burt Perrin and Richard Manning*. London: IACDI. ([www.alnap.org/resource/8496.aspx](http://www.alnap.org/resource/8496.aspx)).

**Riessman, C. K. (1977)** *The effect of interviewer status and respondent sex on symptom reporting*. ([www.alnap.org/resource/8064.aspx](http://www.alnap.org/resource/8064.aspx)).

- Rist, R. C., Boily, M.-H. and Martin, F. (eds) (2011)** *Influencing change: building evaluation capacity to strengthen governance*. Washington, DC: World Bank. ([www.alnap.org/resource/23065.aspx](http://www.alnap.org/resource/23065.aspx)).
- Ritter, T. and Estoesta, J. (2015)** *Ethical challenges in the evaluation of a Pacific multi-country training project by an Australian NGO*. Deakin, ACT: ACFID. ([www.alnap.org/resource/23066.aspx](http://www.alnap.org/resource/23066.aspx)).
- Robert, S. and Valon, A. (2014)** *Quality and Accountability for Project Cycle Management*.
- Roche, C. (1999)** *Impact assessment for development agencies: Learning to value change*. Oxford: Oxfam. ([www.alnap.org/resource/8065.aspx](http://www.alnap.org/resource/8065.aspx)).
- Roche, C. (2000)** *Impact assessment: Seeing the wood and the trees*. Oxford: Oxfam. ([www.alnap.org/resource/11135.aspx](http://www.alnap.org/resource/11135.aspx)).
- Rogers, P. (2012)** *Introduction to impact evaluation*. Washington, DC: InterAction. ([www.alnap.org/resource/6387.aspx](http://www.alnap.org/resource/6387.aspx)).
- Rogers, P. (2013)** *52 weeks of BetterEvaluation: Week 11: Using rubrics*. Melbourne: Better Evaluation. ([www.alnap.org/resource/23067.aspx](http://www.alnap.org/resource/23067.aspx)).
- Rogers, P. (2014a)** *Overview of impact evaluation*. Florence: UNICEF. ([www.alnap.org/resource/22671.aspx](http://www.alnap.org/resource/22671.aspx)).
- Rogers, P. (2014b)** *Overview: Strategies for causal attribution*. Florence: UNICEF. ([www.alnap.org/resource/22674.aspx](http://www.alnap.org/resource/22674.aspx)).
- Rogers, P. (2014c)** *Theory of change*. Florence: UNICEF. ([www.alnap.org/resource/22673.aspx](http://www.alnap.org/resource/22673.aspx)).
- Rosenberg, L. J., Posner, L. D. and Hanley, E. J. (1970)** *Project evaluation and the project appraisal reporting system: Volume one: Summary*. Washington: USAID. ([www.alnap.org/resource/23068.aspx](http://www.alnap.org/resource/23068.aspx)).
- Rouse, A. (2012)** *After action review (AAR) report*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8495.aspx](http://www.alnap.org/resource/8495.aspx)).



**Ryan, M. E. (2009)** 'Making visible the coding process: Using qualitative data software in a post-structural study'. *Issues in Educational Research*, 19(2): 142–161. ([www.alnap.org/resource/23069.aspx](http://www.alnap.org/resource/23069.aspx)).

**Sackett, D. L. (1979)** 'Bias in analytic research'. *Journal of Chronic Diseases*, 32(1–2): 51–63. ([www.alnap.org/resource/23070.aspx](http://www.alnap.org/resource/23070.aspx)).

**Sackett, D. L. (2007)** 'Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't?' *International Journal of Epidemiology*, 36(3): 664–665. ([www.alnap.org/resource/23071.aspx](http://www.alnap.org/resource/23071.aspx)).

**Salama, P., Assefa, F., Talley, L., Spiegel, P., van der Veen, A. and Gotway, C. A. (2001)** 'Malnutrition, measles, mortality, and the humanitarian response during a famine in Ethiopia'. *JAMA*, 286(5): 563–571. ([www.alnap.org/resource/23072.aspx](http://www.alnap.org/resource/23072.aspx)).

**Sandison, P. (2006)** 'The utilisation of evaluations', in *ALNAP review of humanitarian action in 2005: Evaluation utilisation*. ALNAP Review. London: ALNAP. ([www.alnap.org/resource/5225.aspx](http://www.alnap.org/resource/5225.aspx)).

**Sandström, S. and Tchatchua, L. (2010)** 'Do cash transfers improve food security in emergencies? Evidence from Sri Lanka', in Omamo, S. W., Gentilini, U. and Sandström, S. (eds), *Revolution: From food aid to food assistance: Innovations in overcoming hunger*. Rome: WFP. ([www.alnap.org/resource/8077.aspx](http://www.alnap.org/resource/8077.aspx)).

**Saunders, M. (2011)** *Evaluation use and usability: theoretical foundations and current state of international thinking*. 8 July, Gdansk. ([www.alnap.org/resource/23073.aspx](http://www.alnap.org/resource/23073.aspx)).

**Savage, K., Delesgues, L., Martin, E. and Ulfat, G. P. (2007)** *Corruption perceptions and risks in humanitarian assistance: An Afghanistan case study*. London: HPG/ODI. ([www.alnap.org/resource/22801.aspx](http://www.alnap.org/resource/22801.aspx)).

**Sawada, Y. and Shimizutani, S. (2008)** 'How do people cope with natural disasters? Evidence from the great Hanshin-Awaji (Kobe) earthquake in 1995'. *Journal of Money, Credit and Banking*, 40(2-3): 463–488. ([www.alnap.org/resource/23074.aspx](http://www.alnap.org/resource/23074.aspx)).

**Schmidt, W.-P., Arnold, B. F., Boisson, S., Genser, B., Luby, S. P., Barreto, M. L., Clasen, T. and Cairncross, S. (2011)** 'Epidemiological methods in diarrhoea studies—an update'. *International Journal of Epidemiology*, 40(6): 1678–1692. ([www.alnap.org/resource/23075.aspx](http://www.alnap.org/resource/23075.aspx)).

**Schulz, K. F., Chalmers, I., Hayes, R. J. and Altman, D. G. (1995)** 'Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials'. *JAMA*, 273(5): 408–412. ([www.alnap.org/resource/23076.aspx](http://www.alnap.org/resource/23076.aspx)).

**Schwartz, L., Sinding, C., Hunt, M., Elit, L., Redwood-Campbell, L., Adelson, N., Luther, L. Ranford, J. and DeLaat, S. (2010)** 'Ethics in humanitarian aid work: Learning from the narratives of humanitarian health workers'. *AJOB Primary Research*, 1(3): 45–54. ([www.alnap.org/resource/23077.aspx](http://www.alnap.org/resource/23077.aspx)).

**Scriven, M. (1991)** *Evaluation thesaurus*. Newbury Park, CA: SAGE. ([www.alnap.org/resource/8480.aspx](http://www.alnap.org/resource/8480.aspx)).

**Scriven, M. (2008)** 'A summative evaluation of RCT methodology: and an alternative approach to causal research'. *Journal of MultiDisciplinary Evaluation*, 5(9): 14. ([www.alnap.org/resource/8153.aspx](http://www.alnap.org/resource/8153.aspx)).

**Scriven, M. (2011)** *Evaluating evaluations: a meta-evaluation checklist*. ([www.alnap.org/resource/8154.aspx](http://www.alnap.org/resource/8154.aspx)).

**Seebregts, C. J., Zwarenstein, M., Mathews, C., Fairall, L., Flisher, A. J., Seebregts, C., Mukoma, W. and Klepp, K. I. (2009)** 'Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a Palm Pilot interviewing system'. *International journal of medical informatics*. ([www.alnap.org/resource/8155.aspx](http://www.alnap.org/resource/8155.aspx)).

**Serrat, O. (2009)** *The most significant change technique*. Manila: Asian Development Bank. ([www.alnap.org/resource/22802.aspx](http://www.alnap.org/resource/22802.aspx)).

**Shackman, G. (2001)** *Sample size and design effect*, Albany: American Statistical Association. ([www.alnap.org/resource/23080.aspx](http://www.alnap.org/resource/23080.aspx)).

**Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002)** 'Experiments and generalized causal inference and a critical assessment of our assumptions', in *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company. ([www.alnap.org/resource/22889.aspx](http://www.alnap.org/resource/22889.aspx)).

**Shah, R. (2014)** *Evaluation of the Norwegian Refugee Council's Palestine education programme 2010-2014*. Oslo: NRC. ([www.alnap.org/resource/12820.aspx](http://www.alnap.org/resource/12820.aspx)).

**Shanks, L., Ariti, C., Siddiqui, M. R., Pintaldi, G., Venis, S. de Jong, K. and Denault, M. (2013)** 'Counselling in humanitarian settings: a retrospective analysis of 18 individual-focused non-specialised counselling programmes'. *Conflict and Health*, 7(19). ([www.alnap.org/resource/23081.aspx](http://www.alnap.org/resource/23081.aspx)).

**Shannon, H. S., Hutson, R., Kolbe, A., Stringer, B., Haines, T., Nelson, N., Yang, L., Reilly, A., Hardin, J. and Hartley, D. (2012)** 'Choosing a survey sample when data on the population are limited'. *Emerging Themes in Epidemiology*. ([www.alnap.org/resource/8156.aspx](http://www.alnap.org/resource/8156.aspx)).

**Shaw, I., Greene, J. C. and Mark, M. M. (2006)** *The SAGE handbook of evaluation: policies, programs and practices*. London: SAGE. ([www.alnap.org/resource/23082.aspx](http://www.alnap.org/resource/23082.aspx)).

**Shirima, K., Mukasa, O., Armstrong Schellenberg, J., Manzi, F., John, D., Mushi, A., Mrisho, M., Tanner, M., Mshinda, H. and Schellenberg, D. (2007)** 'The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania'. *Emerging themes in epidemiology 2007*. ([www.alnap.org/resource/8157.aspx](http://www.alnap.org/resource/8157.aspx)).

**SIDA. (2011)** *Strategic Evaluation Plan 2012*. Stockholm: SIDA. ([www.alnap.org/resource/23083.aspx](http://www.alnap.org/resource/23083.aspx)).

**Sida, L. (2013)** *Syria crisis: Evaluators learning exchange on 13 September 2013*. London: ALNAP. ([www.alnap.org/resource/9300.aspx](http://www.alnap.org/resource/9300.aspx)).

**Sida, L. (2014)** *Remote monitoring, evaluation and accountability in the Syria response*. London: ALNAP. ([www.alnap.org/resource/12803.aspx](http://www.alnap.org/resource/12803.aspx)).

**Sida, L., Gray, B. and Asmare, E. (2012)** *IASC real time evaluation (IASC RTE) of the humanitarian response to the Horn of Africa drought crisis – Ethiopia*. Geneva: IASC. ([www.alnap.org/resource/7505.aspx](http://www.alnap.org/resource/7505.aspx)).

**Silver, C. and Lewins, A. (2014)** *Using software in qualitative research: a step by step guide*. 2nd ed. Los Angeles, CA: SAGE. ([www.alnap.org/resource/23084.aspx](http://www.alnap.org/resource/23084.aspx)).

**Slim, H. and Trombetta, L. (2014)** *Syria crisis common context analysis*, New York: OCHA. ([www.alnap.org/resource/12718.aspx](http://www.alnap.org/resource/12718.aspx)).

**DANIDA. (2015)** *Evaluation of the Danish engagement in Palestine*. Copenhagen: DANIDA. ([www.alnap.org/resource/23085.aspx](http://www.alnap.org/resource/23085.aspx)).

**SLRC. (2012)** *Making systematic reviews work for international development*. London: SLRC/ODI. ([www.alnap.org/resource/2079.aspx](http://www.alnap.org/resource/2079.aspx)).

**Small, M .L. (2011)** 'How to conduct a mixed methods study: Recent trends in a rapidly growing literature'. *Annual Review of Sociology*, 37(1): 57–86. ([www.alnap.org/resource/23086.aspx](http://www.alnap.org/resource/23086.aspx)).

**SMART, 2012.** *Sampling methods and sample size calculation for the SMART methodology*. Toronto: SMART. ([www.alnap.org/resource/23087.aspx](http://www.alnap.org/resource/23087.aspx)).

**Smutylo, T. (2001)** *Crouching impact, hidden attribution: Overcoming threats to learning in development programs*. Ottawa: International Development Research Centre. ([www.alnap.org/resource/19253.aspx](http://www.alnap.org/resource/19253.aspx)).

**Sphere Project. (2011)** *Humanitarian charter and minimum standards in humanitarian response*. Geneva: Sphere Project. ([www.alnap.org/resource/8161.aspx](http://www.alnap.org/resource/8161.aspx)).

**Start Fund. (2014)** *South Sudan crisis response summary dashboard crisis response*. London: Start Fund. ([www.alnap.org/resource/19411.aspx](http://www.alnap.org/resource/19411.aspx)).

**Start, D. and Hovland, I. (2004)** *Tools for policy impact: A handbook for researchers*. London: ODI. ([www.alnap.org/resource/8479.aspx](http://www.alnap.org/resource/8479.aspx)).

**Steets, J., Darcy, J., Weingartner, L. and Leguene, P. (2014)** *Strategic evaluation FAO/WFP joint evaluation of food security cluster coordination in humanitarian action*. Rome: WFP. ([www.alnap.org/resource/19699.aspx](http://www.alnap.org/resource/19699.aspx)).

**Steets, J. and Dubai, K. (2010)** *Real-time evaluation of UNICEF's response to the Sa'ada conflict in Northern Yemen*. Berlin: GPPI. ([www.alnap.org/resource/8896.aspx](http://www.alnap.org/resource/8896.aspx)).

**Steets, J., Grünewald, F., Binder, A., de Geoffroy, V., Kauffmann, D., Krüger, S., Meier, C., and Sokpoh, B. (2010)** *Cluster approach evaluation 2 Synthesis Report*. Berlin: GPPi. ([www.alnap.org/resource/5986.aspx](http://www.alnap.org/resource/5986.aspx)).

**Stein, D. and Valters, C. (2012)** *Understanding theory of change in international development: a review of existing knowledge*. London: London School of Economics. ([www.alnap.org/resource/23088.aspx](http://www.alnap.org/resource/23088.aspx)).

**Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., and Befani, B. (2012)** *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*. London: DFID. ([www.alnap.org/resource/8196.aspx](http://www.alnap.org/resource/8196.aspx)).

**Stetson, V. (2008)** *Communicating and reporting on an evaluation. Guidelines and tools*. Washington, DC: American Red Cross. ([www.alnap.org/resource/9888.aspx](http://www.alnap.org/resource/9888.aspx)).

**Stoddard, A., Harmer, A., Haver, K., Salomons, D. and Wheeler, V. (2007)** *Cluster approach evaluation*. New York: OCHA. ([www.alnap.org/resource/3641.aspx](http://www.alnap.org/resource/3641.aspx)).

**Stokke, K. (2007)** *Humanitarian response to natural disasters: a synthesis of evaluation findings*. Oslo: NORAD. ([www.alnap.org/resource/3551.aspx](http://www.alnap.org/resource/3551.aspx)).

**Street, A. (2009)** *Synthesis report: Review of the engagement of NGOs with the humanitarian reform process*. ([www.alnap.org/resource/12647.aspx](http://www.alnap.org/resource/12647.aspx)).

**Suchman, L. and Jordan, B. (1990)** 'Interactional troubles in face-to-face survey interviews'. *Journal of the American Statistical Association*, 85(409): 232–241. ([www.alnap.org/resource/23089.aspx](http://www.alnap.org/resource/23089.aspx)).

**Sutter, P., Downen, J., Walters, T., Izabiriza, B., Kagendo, R. S., Langworthy, M., Sagara, B. and Mueller, M. (2012)** *The contribution of food assistance to Durable Solutions in protracted refugee situations: its impact and role in Rwanda (2007–2011)*. Rome: WFP. ([www.alnap.org/resource/19832.aspx](http://www.alnap.org/resource/19832.aspx)).

**TANGO International. (2007)** *End of project study of tsunami impacted communities in Southern India*. Tucson, AZ: TANGO International. ([www.alnap.org/resource/3536.aspx](http://www.alnap.org/resource/3536.aspx)).

**Taplin, D. H. and Heléne, C. (2012)** *Theory of change basics: A primer on theory of change*. New York: ActKnowledge. ([www.alnap.org/resource/22902.aspx](http://www.alnap.org/resource/22902.aspx)).

**Taplin, D. H., Clark, H., Collins, E. and Colby, D. C. (2013)** *Theory of change: Technical papers: a series of papers to support development of theories of change based on practice in the field*. New York: ActKnowledge. ([www.alnap.org/resource/23090.aspx](http://www.alnap.org/resource/23090.aspx)).

**Taylor-Powell, E. and Henert, E. (2008)** *Developing a logic model: Teaching and training guide*. Madison, WI: University of Wisconsin. ([www.alnap.org/resource/23091.aspx](http://www.alnap.org/resource/23091.aspx)).

**Tearfund. (2015)** *Impact and learning report 2015: Inspiring change*. Teddington: Tearfund. ([www.alnap.org/resource/23092.aspx](http://www.alnap.org/resource/23092.aspx)).

**Teddlie, C. and Yu, F. (2007)** 'Mixed methods sampling: A typology with examples'. *Journal of Mixed Methods Research*, 1(1): 77–100. ([www.alnap.org/resource/23093.aspx](http://www.alnap.org/resource/23093.aspx)).

**Telford, J. (1997)** *Counting and identification of beneficiary populations in emergencies: registration and its alternatives*. London: ODI. ([www.alnap.org/resource/11010.aspx](http://www.alnap.org/resource/11010.aspx)).

**Telford, J. (2009)** *Review of joint evaluations and the future of inter agency evaluations*. New York: OCHA. ([www.alnap.org/resource/5772.aspx](http://www.alnap.org/resource/5772.aspx)).

**Telford, J. and Cosgrave, J. (2007)** *The international humanitarian system and the 2004 Indian Ocean earthquake and tsunamis*. ([www.alnap.org/resource/8162.aspx](http://www.alnap.org/resource/8162.aspx)).

**Telford, J., Cosgrave, J. and Houghton, R. (2006)** *Joint evaluation of the international response to the Indian Ocean tsunami: Synthesis report*. London: TEC. ([www.alnap.org/resource/3535.aspx](http://www.alnap.org/resource/3535.aspx)).

**Telyukov, A. and Paterson, M. (2008)** *Impact evaluation of PRM humanitarian assistance to the repatriation and reintegration of Burundi refugees (2003-08)*. Washington, DC: USAID. ([www.alnap.org/resource/5661.aspx](http://www.alnap.org/resource/5661.aspx)).

**Ternström, B. (2013)** *Evaluation of five humanitarian programmes of the Norwegian Refugee Council and of the standby roster: NORCAP*. Oslo: NORAD. ([www.alnap.org/resource/10101.aspx](http://www.alnap.org/resource/10101.aspx)).

**Tessitore, S. (2013)** "Like a good trip to town without selling your animals": *A study of FAO Somalia's cash for work programme*. Rome: FAO. ([www.alnap.org/resource/8821.aspx](http://www.alnap.org/resource/8821.aspx)).

**Theis, J. and Grady, H. (1991)** *Participatory rapid appraisal for community development: A training manual based on experiences in the Middle East and North Africa*. London: IIED. ([www.alnap.org/resource/8163.aspx](http://www.alnap.org/resource/8163.aspx)).

- Thizy, D. (2013)** *Federation-wide livelihood program evaluation*. Geneva: IFRC. ([www.alnap.org/resource/8864.aspx](http://www.alnap.org/resource/8864.aspx)).
- Thomas, A. and Mohnan, G. (2007)** *Research skills for policy and development: how to find out fast*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8164.aspx](http://www.alnap.org/resource/8164.aspx)).
- Thomas, V. and Beck, T. (2010)** *Changing the way UNHCR does business? An evaluation of the age, gender and diversity mainstreaming strategy, 2004-2009*. Geneva: UNHCR. ([www.alnap.org/resource/5852.aspx](http://www.alnap.org/resource/5852.aspx)).
- Todd, D., Batchelor, C., Brouder, S., Coccia, F., Castiel, E. F. and Soussan, J. (2015)**  
Evaluation of the CGIAR research program on water, land and ecosystems. Montpellier: CGIAR. ([www.alnap.org/resource/22803.aspx](http://www.alnap.org/resource/22803.aspx)).
- Tolani-Brown, N., Mortensen, J. and Davis, J. (2010)** *Evaluation of the UNICEF education programme in Timor Leste 2003 – 2009*. New York: UNICEF. ([www.alnap.org/resource/6383.aspx](http://www.alnap.org/resource/6383.aspx)).
- Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., Tsai, A. and Jackson, D. (2009)** *The use of mobile phones as a data collection tool: A report from a household survey in South Africa*. London: BioMed Central. ([www.alnap.org/resource/8165.aspx](http://www.alnap.org/resource/8165.aspx)).
- Beck, T. (2011)** *Joint humanitarian impact evaluation: Report for the inter-agency working group on joint humanitarian impact evaluation*. London: TEC. ([www.alnap.org/resource/6046.aspx](http://www.alnap.org/resource/6046.aspx)).
- tools4dev. (2014)** *How to pretest and pilot a survey questionnaire*. ([www.alnap.org/resource/22804.aspx](http://www.alnap.org/resource/22804.aspx)).
- Tufte, E. R. (1990)** *Envisioning information*. Graphics Press. ([www.alnap.org/resource/8166.aspx](http://www.alnap.org/resource/8166.aspx)).
- Tufte, E. R. (1997)** *Visual explanations: images and quantities, evidence and narrative*. Graphics Press. ([www.alnap.org/resource/22805.aspx](http://www.alnap.org/resource/22805.aspx)).
- Tufte, E. R. and Graves-Morris, P. R. (2001)** *The visual display of quantitative information*. Graphics Press. ([www.alnap.org/resource/8167.aspx](http://www.alnap.org/resource/8167.aspx)).
- Tulane University. (2012)** *Haiti humanitarian assistance evaluation from a resilience perspective*. New Orleans, LA: Tulane University. ([www.alnap.org/resource/6377.aspx](http://www.alnap.org/resource/6377.aspx)).
- Tuttle, A. H., Tohyama, S., Ramsay, T., Kimmelman, J., Schweinhardt, P, Bennett, G., Mogil, J. (2015)** 'Increasing placebo responses over time in U.S. clinical trials of neuropathic pain'. *Pain*, 156(12): 2616-2626. ([www.alnap.org/resource/22806.aspx](http://www.alnap.org/resource/22806.aspx)).

**Ubels, J., Acquaye-Baddoo, N. A. and Fowler, A. (2010)** *Capacity development in practice*. London: Earthscan. ([www.alnap.org/resource/22808.aspx](http://www.alnap.org/resource/22808.aspx)).

**UN General Assembly. (1989)** *Convention on the rights of the child*. New York: UN General Assembly. ([www.alnap.org/resource/22809.aspx](http://www.alnap.org/resource/22809.aspx)).

**UN General Assembly. (1951)** *Convention relating to the status of refugees*. New York: UN General Assembly. ([www.alnap.org/resource/22810.aspx](http://www.alnap.org/resource/22810.aspx)).

**UNDP. (2009)** *Handbook on planning, monitoring and evaluating for development results*. New York: UNDP. ([www.alnap.org/resource/22812.aspx](http://www.alnap.org/resource/22812.aspx)).

**UNDP. (2011)** *Updated guidance on evaluation in the handbook on planning, monitoring and evaluation for development results (2009)*. New York: UNDP. ([www.alnap.org/resource/22813.aspx](http://www.alnap.org/resource/22813.aspx)).

**UNEG. (2005a)** *Norms for evaluation in the UN system*. New York: UNEG. ([www.alnap.org/resource/8215.aspx](http://www.alnap.org/resource/8215.aspx)).

**UNEG. (2005b)** *Standards for evaluation in the UN system*. New York: UNEG. ([www.alnap.org/resource/19410.aspx](http://www.alnap.org/resource/19410.aspx)).

**UNEG. (2008a)** *Core competencies for evaluators of the UN System*, New York: UNEG. ([www.alnap.org/resource/22814.aspx](http://www.alnap.org/resource/22814.aspx)).

**UNEG. (2008b)** *Evaluability assessments of the programme country pilots: Delivering as one UN*. New York: UNEG. ([www.alnap.org/resource/22815.aspx](http://www.alnap.org/resource/22815.aspx)).

**UNEG. (2010a)**. *Good practice guidelines for follow up to evaluations*. New York: UNEG. ([www.alnap.org/resource/8446.aspx](http://www.alnap.org/resource/8446.aspx)).

**UNEG. (2010b)** *UNEG quality checklist for evaluation reports*. New York: UNEG. ([www.alnap.org/resource/22816.aspx](http://www.alnap.org/resource/22816.aspx)).

**UNEG. (2011)** *Integrating human rights and gender equality in evaluation - Towards UNEG guidance*. New York: UNEG. ([www.alnap.org/resource/8171.aspx](http://www.alnap.org/resource/8171.aspx)).

**UNEG. (2013a)** *Resource pack on joint evaluations*. New York: UNEG. ([www.alnap.org/resource/12613.aspx](http://www.alnap.org/resource/12613.aspx)).

**UNEG. (2013b)** *UNEG handbook for conducting evaluations of normative work in the UN system*. New York: UNEG. ([www.alnap.org/resource/22817.aspx](http://www.alnap.org/resource/22817.aspx)).

**UNEG. (2014)** *Integrating human rights and gender equality in evaluation - Towards UNEG guidance*. New York: UNEG. ([www.alnap.org/resource/19283.aspx](http://www.alnap.org/resource/19283.aspx)).

**UNEG HEIG. (2016)** *Reflecting Humanitarian Principles in Evaluation*. New York: UNEG. (<http://www.alnap.org/resource/23385.aspx>).



**UNHCR. (1997)** *Commodity distribution*. Geneva: UNHCR. ([www.alnap.org/resource/22818.aspx](http://www.alnap.org/resource/22818.aspx)).

**UNHCR. (2007)** *Handbook for emergencies*. New York: UNHCR. ([www.alnap.org/resource/8172.aspx](http://www.alnap.org/resource/8172.aspx)).

**UNHCR and WFP. (2013)** *Synthesis report of the joint WFP and UNHCR impact evaluations on the contribution of food assistance to durable solutions in protracted refugee situations*. Geneva and Rome: UNHCR/WFP. ([www.alnap.org/resource/12260.aspx](http://www.alnap.org/resource/12260.aspx)).

**UNICEF. (1998)** *The state of the world's children*. New York: UNICEF. ([www.alnap.org/resource/19280.aspx](http://www.alnap.org/resource/19280.aspx)).

**UNICEF. (2002)** *Children participating in research, monitoring and evaluation - Ethics and your responsibilities as a manager*. New York: UNICEF. ([www.alnap.org/resource/8174.aspx](http://www.alnap.org/resource/8174.aspx)).

**UNICEF. (2005)** *Emergency field handbook*. New York: UNICEF. ([www.alnap.org/resource/8173.aspx](http://www.alnap.org/resource/8173.aspx)).

**UNICEF. (2010)** *Evaluation of the UNICEF education programme in Timor Leste 2003 – 2009*. New York: UNICEF. ([www.alnap.org/resource/6383.aspx](http://www.alnap.org/resource/6383.aspx)).

**UNICEF. (2013a)** *Evaluability assessment of the peacebuilding, education and advocacy programme*. New York: UNICEF. ([www.alnap.org/resource/12679.aspx](http://www.alnap.org/resource/12679.aspx)).

**UNICEF. (2013b)** *Evaluation of UNICEF programmes to protect children in emergencies: Pakistan country case study*. New York: UNICEF. ([www.alnap.org/resource/22820.aspx](http://www.alnap.org/resource/22820.aspx)).

**UNICEF. (2013c)** *Evaluation of UNICEF programmes to protect children in emergencies. Synthesis report*. New York: UNICEF. ([www.alnap.org/resource/12491.aspx](http://www.alnap.org/resource/12491.aspx)).

**UNICEF. (2013d)** *Global evaluation reports oversight system (GEROS)*. New York: UNICEF. ([www.alnap.org/resource/7948.aspx](http://www.alnap.org/resource/7948.aspx)).

**UNICEF. (2013e)** *Thematic synthesis report on evaluation of humanitarian action*. New York: UNICEF. ([www.alnap.org/resource/8442.aspx](http://www.alnap.org/resource/8442.aspx)).

**UNICEF. (2013f)** *Improving Child Nutrition: The achievable imperative for global progress*. New York: UNICEF. (<http://www.alnap.org/resource/23386.aspx>).

**UNICEF. (2014)** *Evaluation of UNICEF's cluster lead agency role in humanitarian action (CLARE)*. New York: UNICEF. ([www.alnap.org/resource/19327.aspx](http://www.alnap.org/resource/19327.aspx)).

**UNICEF. (2015)** *Final Report for the formative evaluation of the highscope curriculum reform programme (February to December 2014)*. Christ Church: UNICEF. ([www.alnap.org/resource/22819.aspx](http://www.alnap.org/resource/22819.aspx)).

**UNICEF and Save the Children. (2010)** *Management response to the review of the global education cluster co-leadership arrangement*. New York and London: UNICEF/Save the Children. ([www.alnap.org/resource/6354.aspx](http://www.alnap.org/resource/6354.aspx)).

**UNIFEM. (2011)** *UNIFEM strategic plan 2008-2011: Evaluability assessment*. New York: UNIFEM. ([www.alnap.org/resource/22821.aspx](http://www.alnap.org/resource/22821.aspx)).

**Universalia. (2013)** *GEROS – Global meta-evaluation report 2012: final report*, New York: UNICEF. ([www.alnap.org/resource/22822.aspx](http://www.alnap.org/resource/22822.aspx)).

**UNODC. (2012)** *Evaluability assessment template*. Vienna: UNODC. ([www.alnap.org/resource/22823.aspx](http://www.alnap.org/resource/22823.aspx)).

**USAID. (1996)** *Preparing an evaluation scope of work*. Washington: USAID. ([www.alnap.org/resource/22825.aspx](http://www.alnap.org/resource/22825.aspx)).

**USAID. (2010)** *Performance monitoring and evaluation tips using rapid appraisal methods*. Washington, DC: USAID. ([www.alnap.org/resource/19281.aspx](http://www.alnap.org/resource/19281.aspx)).

**USAID, 2012.** *How-to note: Preparing evaluation reports*, Washington, DC: USAID. ([www.alnap.org/resource/22824.aspx](http://www.alnap.org/resource/22824.aspx)).

**USAID. (2013)** *After-action review guidance*. Washington, DC: USAID. ([www.alnap.org/resource/19257.aspx](http://www.alnap.org/resource/19257.aspx)).

**USAID. (2015)** *USAID office of food for peace food security desk review for Katanga, North Kivu and South Kivu, Democratic Republic of Congo*. Washington, DC: USAID. ([www.alnap.org/resource/22826.aspx](http://www.alnap.org/resource/22826.aspx)).

**Valters, C. (2014)** *Theories of change in international development: Communication, learning, or accountability?* London: London School of Economics. ([www.alnap.org/resource/19265.aspx](http://www.alnap.org/resource/19265.aspx)).

**Van Bruaene, M., Dumélie, R., Kunze, M., Pankhurst, M. and Potter, J. (2010)** *Review concerning the establishment of a European voluntary humanitarian aid corps*. Brussels: Prolog Consult. ([www.alnap.org/resource/3633.aspx](http://www.alnap.org/resource/3633.aspx)).

**Varanoa, S. P., Schaferb, J. A., Cancinoc, J. M., Deckerd, S., H. and Greenee, J. R. (2010)** 'A tale of three cities: Crime and displacement after Hurricane Katrina'. *Journal of Criminal Justice*, 38(1): 42-50. ([www.alnap.org/resource/22827.aspx](http://www.alnap.org/resource/22827.aspx)).

**Wagner, A. K., Soumerai, S. B., Zhang, F. and Ross-Degnan, D. (2002)** 'Segmented regression analysis of interrupted time series studies in medication use research'. *Journal of Clinical Pharmacy and Therapeutics*, 2002(27): 299-309. ([www.alnap.org/resource/22830.aspx](http://www.alnap.org/resource/22830.aspx)).

**Walden, V. (2013)** *A quick guide to monitoring, evaluation, accountability, and learning in fragile contexts*. Oxford: Oxfam. ([www.alnap.org/resource/10593.aspx](http://www.alnap.org/resource/10593.aspx)).

**Warner, A. (2014)** *EHA, let's do some stretches! ALNAP EHA Practice Note*. Shared in ALNAP Humanitarian Evaluation Community of Practice in October.

**Warner, A. (2014)** *Evaluation of humanitarian action discussion series. Repeat after me: communicate, disseminate and support take-up!* London: ALNAP. ([www.alnap.org/resource/12920.aspx](http://www.alnap.org/resource/12920.aspx)).

**Webb, E. J., Campbell, D. T., Schwartz, R. D. and Sechrest, L. (1966)** *Unobtrusive measures: nonreactive research in the social sciences*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22831.aspx](http://www.alnap.org/resource/22831.aspx)).

**Webster, C. (1996)** 'Hispanic and Anglo interviewer and respondent ethnicity and gender: The impact on survey response quality. *Journal of Marketing Research*, 33(1): 62–72. ([www.alnap.org/resource/8176.aspx](http://www.alnap.org/resource/8176.aspx)).

**Weingärtner, L., Otto, R. and Hoerz, T. (2011)** *Die deutsche humanitäre Hilfe im Ausland. Band I: Hauptbericht*. Bonn: BMZ. ([www.alnap.org/resource/22832.aspx](http://www.alnap.org/resource/22832.aspx)).

**Weiss, C. H. (1997)** *Evaluation methods for studying programs and policies*. Upper Saddle River, NJ: Prentice Hall. ([www.alnap.org/resource/8540.aspx](http://www.alnap.org/resource/8540.aspx)).

**Wesonga, D. (2013)** *Accelerated primary education support (APES) project: Final evaluation report*. London: Concern Worldwide. ([www.alnap.org/resource/12232.aspx](http://www.alnap.org/resource/12232.aspx)).

**WFP. (2002)** *Emergency field operations pocketbook*. Rome: WFP. ([www.alnap.org/resource/8177.aspx](http://www.alnap.org/resource/8177.aspx)).

**WFP. (2010)** *Evaluation Matrix*. Rome: WFP. ([www.alnap.org/resource/22833.aspx](http://www.alnap.org/resource/22833.aspx)).

**WFP. (2012)** *Mixed-method impact evaluation - The contribution of food assistance to durable solutions in protracted refugee situations: its impact and role in Bangladesh*. Rome: WFP. ([www.alnap.org/resource/19830.aspx](http://www.alnap.org/resource/19830.aspx)).

**WFP. (2013)** *Synthesis of the series of joint UNHCR-WFP impact evaluations of food assistance to refugees in protracted situations*. Rome: WFP. ([www.alnap.org/resource/19903](http://www.alnap.org/resource/19903)).

**WFP. (2014)** *Impact evaluations I. guidance for process and content*. Rome: WFP. ([www.alnap.org/resource/22834.aspx](http://www.alnap.org/resource/22834.aspx)).

**WFP. (2015)** *Evaluation quality assurance system (EQAS) guidelines for operation evaluations*. Rome: WFP. ([www.alnap.org/resource/22835.aspx](http://www.alnap.org/resource/22835.aspx)).

- White, H. (2009a)** *Some reflections on current debates in impact evaluation*. New Delhi: 3ie. ([www.alnap.org/resource/8178.aspx](http://www.alnap.org/resource/8178.aspx)).
- White, H. (2009b)** *Theory-based impact evaluation: principles and practice*. New Delhi: 3ie. ([www.alnap.org/resource/8179.aspx](http://www.alnap.org/resource/8179.aspx)).
- White, H., 2014.** *The counterfactual. Howard White International Initiative for Impact Evaluation*. New Delhi: 3ie. ([www.alnap.org/resource/22836.aspx](http://www.alnap.org/resource/22836.aspx)).
- White, H. and Sabarwal, S. (2014a)** *Developing and selecting measures of child well-being*. Florence: UNICEF. ([www.alnap.org/resource/22837.aspx](http://www.alnap.org/resource/22837.aspx)).
- White, H. and Sabarwal, S. (2014b)** *Modelling*. Florence: UNICEF. ([www.alnap.org/resource/22838.aspx](http://www.alnap.org/resource/22838.aspx)).
- White, H. and Sabarwal, S. (2014c)** *Quasi-experimental design and methods*. Florence: UNICEF. ([www.alnap.org/resource/22839.aspx](http://www.alnap.org/resource/22839.aspx)).
- White, H., Sabarwal, S. and de Hoop, T. (2014)** *Randomized controlled trials (RCTs)*. Florence: UNICEF. ([www.alnap.org/resource/22840.aspx](http://www.alnap.org/resource/22840.aspx)).
- van der Wijk, J. and Kazan, L. (2010)** *Evaluation of the DEC-funded CAFOD health and WASH project in the DRC*. London: CAFOD. ([www.alnap.org/resource/5849.aspx](http://www.alnap.org/resource/5849.aspx)).
- Wilding, J., Swift, J. and Hartung, H. (2009)** *Mid term evaluation of DG ECHO's regional drought decision in the Greater Horn of Africa: March - May 2009*. Brussels: ECHO. ([www.alnap.org/resource/5919.aspx](http://www.alnap.org/resource/5919.aspx)).
- Wiles, P., Chan, L. and Horwood, C. (1999)** *Evaluation of Danish humanitarian assistance to Afghanistan 1992-98*. Copenhagen: DANIDA. ([www.alnap.org/resource/2827.aspx](http://www.alnap.org/resource/2827.aspx)).
- Willitts-King, B, Taylor, G. and Barber, K. (2012)** *Process Review of the Common Humanitarian Fund for Somalia*. New York: OCHA (<http://www.alnap.org/resource/6340>).
- Wilson, P. and Reilly, D. (2007)** *Joint evaluation of their responses to the Yogyakarta earthquake July 2007*. Atlanta, GA: CARE International. ([www.alnap.org/resource/3611.aspx](http://www.alnap.org/resource/3611.aspx)).
- W.K. Kellogg Foundation. (1998)** *W.K. Kellogg Foundation evaluation handbook*. Albuquerque, NM: ([www.alnap.org/resource/22841.aspx](http://www.alnap.org/resource/22841.aspx)).
- World Bank. (2004)** *Citizen report card surveys - A note on the concept and methodology*. Washington, DC: World Bank. ([www.alnap.org/resource/8182.aspx](http://www.alnap.org/resource/8182.aspx)).

**World Bank. (2009)** *Making smart policy: using impact evaluation for policy making*. Washington, DC: World Bank. ([www.alnap.org/resource/8434.aspx](http://www.alnap.org/resource/8434.aspx)).

**World Bank. (2007a)** 'Participatory tools for micro-level poverty and social impact analysis', in *Tools for institutional, political, and social analysis of policy reform*. Washington, DC: World Bank. ([www.alnap.org/resource/22842.aspx](http://www.alnap.org/resource/22842.aspx)).

**World Bank. (2007b)** *Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards*, Washington: The Independent Evaluation Group of the World Bank. ([www.alnap.org/resource/22843.aspx](http://www.alnap.org/resource/22843.aspx)).

**Wortel, E. (2009)** *Humanitarians and their moral stance in war: the underlying values*. Geneva: ICRC. ([www.alnap.org/resource/8183.aspx](http://www.alnap.org/resource/8183.aspx)).

**Yang, C. H. Xirasagar, S., Chung, H. C., Huang, Y. T. and Lin, H. C. (2005)** 'Suicide trends following the Taiwan earthquake of 1999: empirical evidence and policy implications'. *Acta Psychiatrica Scandinavica*, 112: 442–448. ([www.alnap.org/resource/8184.aspx](http://www.alnap.org/resource/8184.aspx)).

**Yarborough, D., Shulha, L., Hopson, R. and Caruthers, F.** *The program evaluation standards*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22725.aspx](http://www.alnap.org/resource/22725.aspx)).

**Zaiiontz, C. (2015)** *Performing real statistical analysis using excel*. ([www.alnap.org/resource/22844.aspx](http://www.alnap.org/resource/22844.aspx)).