

INSUFFISANCE DES PREUVES ?

Qualité et exploitation des données probantes dans l'action humanitaire

Paul Knox Clarke et James Darcy

ALNAP est un système sectoriel unique dont la mission est d'améliorer la performance de l'action humanitaire par l'apprentissage partagé.

www.alnap.org

Avez-vous lu **l'amorce de discussion ALNAP** qui accompagne cette étude ?

www.alnap.org/pool/files/discussion-starter-evidence-alnap-2014.pdf

Une copie électronique de cette étude, l'amorce de discussion et autres ressources connexes sont disponibles sur le site web ALNAP à www.alnap.org/ourwork/evidence.

Suggestion de référencement bibliographique :

Knox Clarke, P. et Darcy, J. (2014) *Insuffisance des preuves ? Qualité et exploitation des données probantes dans l'action humanitaire*. Londres : ALNAP/ODI.

© ALNAP/ODI 2014. Ces travaux sont sous couverts par une licence Commons Attribution – licence non commerciale (CC BY-NC 3.0).

ISBN 978-1-909464-71-1

Publication et communication gérées par Franziska Orphal

Révision de la copie par Angela Hawke

Traduction par Phillipe Galinier

Couverture et composition par Soapbox



Sommaire

Remerciements	3
Introduction	5
1.1 Importance des données probantes pour l'action humanitaire	5
1.2 Plan de la présente étude	7
Qu'entend-on par « données probantes » dans le contexte de l'action humanitaire et comment en juger la qualité ?	8
2.1 Contexte : termes, concepts et défis	8
2.2 Des données probantes pour quoi ? Propositions humanitaires et exigences de données probantes	12
2.3 Critères d'évaluation de la qualité des données probantes humanitaires	15
Les données probantes humanitaires répondent-elles à l'heure actuelle aux critères de qualité ?	20
3.1 Les enjeux de la production de données probantes dans les situations humanitaires	20
3.2 Données probantes émanant des systèmes d'alerte précoce	22
3.3 Données probantes issues de l'évaluation des besoins	25
3.4 Données probantes émanant du suivi	34
3.5 Données probantes émanant des évaluations et des essais contrôlés	42
Les données probantes actuelles sont-elles exploitées pour orienter les opérations et l'élaboration des politiques humanitaires ?	51
4.1 La qualité n'est pas un gage d'utilisation	51
4.2 Exploitation des données probantes émanant de l'alerte précoce	51
4.3 Exploitation des données probantes fournies par l'évaluation et le suivi	55
4.4 Utilisation des données issues des évaluations	60
Conclusions et recommandations : Comment peut-on améliorer la qualité et l'exploitation des données dans l'action humanitaire ?	72
5.1 Améliorer la qualité des données	72
5.2 Améliorer l'utilisation des données probantes dans la prise de décision humanitaire	77
Annexe 1 : Méthodologie	81
Annexe 2 : Liste des personnes interrogées et des contributeurs	82
Bibliographie	83

Remerciements

La présente étude a été préparée par le secrétariat d'ALNAP et s'inscrit dans le prolongement de la communication « Données et savoirs dans l'action humanitaire » présentée à la 28^e réunion annuelle d'ALNAP (mars 2013). L'étude originale a été rédigée par James Darcy et Paul Knox Clarke, avec le concours de Miriam Bradley, Alexandra Warner et Leah Campbell pour les recherches. Les auteurs souhaitent remercier tous ceux qui ont participé à la réunion annuelle, et tout particulièrement les personnes qui ont fait des présentations : leurs expériences et leurs points de vue forment un élément important de cette version révisée. Les auteurs remercient également toutes les personnes, dont les noms figurent en annexe 2, qui ont accepté d'être interrogées dans le cadre de cette étude. Enfin, ils souhaitent exprimer leur gratitude envers les lecteurs critiques Richard Garfield, Joanna Macrae et Anthony Redmond dont les suggestions ont considérablement rehaussé la qualité de cette étude. Les auteurs conservent l'entière responsabilité des opinions exprimées et des erreurs ou omissions éventuelles qui se seraient glissées dans le texte.



Introduction

1.1 Importance des données probantes pour l'action humanitaire

Le manque de production et d'exploitation de données probantes dans l'élaboration de politiques et l'intervention rend l'action humanitaire moins efficace, moins conforme à l'éthique et moins redevable.

L'efficacité des interventions humanitaires dépend autant des savoirs acquis que des financements et de la logistique. En effet, faute de connaissances, les interventions ont peu de chances de réussir : « La capacité de la communauté humanitaire internationale à recueillir, analyser, diffuser et exploiter les informations clés est fondamentale pour l'efficacité de l'intervention » (Mahmood et coll., 2010).

De plus, dans un secteur si ancré dans les valeurs, il est capital de disposer de données probantes si nous voulons défendre les principes humanitaires : « Si nous croyons à l'impartialité, nous devons nous fonder sur des données factuelles. Nous ne pouvons pas être impartiaux si nous ignorons quels choix s'offrent à nous » (Peter Walker, communication personnelle, mars 2013). Une étude récente du DFID sur les défis liés à la collecte de données probantes dans les situations d'urgence souligne la nécessité de suivre des cadres éthiques en la matière et affirme qu'il est « contraire à l'éthique de monter des interventions qui au mieux ne reposent pas sur des données probantes, qui sont inefficaces ou, pire, sont véritablement nuisibles » (DFID, 2012 p. 11).

Enfin, les données probantes sont importantes pour la redevabilité. Les organismes de donateurs, les États sinistrés et la société civile ont le droit de savoir comment – et avec quelle efficacité – l'argent est dépensé au nom des populations touchées par des crises. Pour répondre à ces attentes, les organisations humanitaires doivent pouvoir prouver l'existence de besoins et démontrer qu'elles ont fait des choix informés et délibérés sur la manière la plus efficace et la plus efficiente d'intervenir. Elles doivent aussi pouvoir fournir des données sur l'impact des choix qu'elles ont faits.

La présente étude (et la réunion d'ALNAP sur laquelle elle repose) part du sentiment que, « à présent, les décisions humanitaires se fondent souvent sur des informations médiocres » (DFID, 2012 p. 5) et sont « motivées par l'anecdote plutôt que par des données probantes » (Mazurana et coll., 2011 p. 1). Et même lorsque des éléments probants sont disponibles, il semble parfois que les décisions continuent d'être motivées par la conviction personnelle, ou par des considérations politiques ou liées à la collecte de fonds. S'il est donc important que les données disponibles soient de la plus haute qualité, il est tout aussi important que ces données soient *exploitées* par les décideurs. Les échecs collectifs récurrents à réagir résolument face à des preuves solides de crises imminentes (découlant notamment des systèmes d'alerte



L'efficacité des interventions humanitaires dépend autant des savoirs acquis que des financements et de la logistique.



précoce à la famine en Afrique subsaharienne) confirment bien que la *production* de telles données n'est qu'un aspect des enjeux. Il en va de même pour les éléments factuels émanant d'expériences passées : l'un des thèmes récurrents des évaluations est que le système international et ses organismes individuels ont du mal à tirer les enseignements du passé et à appliquer les données factuelles qui en découlent aux pratiques humanitaires d'aujourd'hui (Sandison, 2006 ; Hallam, 2011).

L'action humanitaire n'est pas le seul domaine confronté à de telles préoccupations. Au cours des 15 dernières années, maints domaines de politique publique – la santé, l'éducation et le développement international en autres – ont reconnu le besoin d'intégrer plus systématiquement les données probantes aux pratiques. Stern et col. (2012) décrivent l'émergence d'un courant stratégique fondé sur les données probantes (Evidence Based Policy ou EBP en anglais), selon lequel les dirigeants devraient prendre leurs décisions en fonction de preuves plutôt que de leur idéologie ou par réaction à des intérêts particuliers (voir également Davies et coll., 2000 ; Nutley et coll., NAO, 2003). Cette plus grande importance accordée aux données probantes a suscité un débat autour du sens même du terme, sur ce qui constitue une « donnée probante » et sur les meilleures démarches pour en produire. La présente étude examine certaines de ces questions dans le contexte spécifique de la réponse internationale aux situations d'urgence humanitaire.

La production et l'exploitation de données probantes dans l'action humanitaire ne sont pas des thèmes nouveaux : dans les années 1990, les discussions sur ces questions faisaient partie des efforts visant à renforcer la qualité et la redevabilité dans l'action humanitaire. Depuis, plusieurs initiatives techniques ont été menées afin d'améliorer la qualité des informations que nous fournissent les mécanismes d'alerte précoce, d'appréciation et d'évaluation. Leur but global est d'assurer que les interventions humanitaires sont réellement fondées sur les besoins et sur des pratiques efficaces avérées.

Des actions visant à assurer une exploitation plus généralisée des données probantes dans l'action humanitaire ont également été tentées. De nombreux organismes ont essayé de renforcer leurs propres systèmes d'apprentissage par l'organisation et de gestion des savoirs, tandis que le système dans son ensemble a soutenu des initiatives collectives en faveur de l'apprentissage commun, de la codification et de la fixation de normes (Walker et Purdin, 2004 ; Young et Harvey, 2004 ; Barnett, 2005). Il reste néanmoins du chemin à parcourir pour améliorer tant la qualité que l'exploitation des données dans le secteur humanitaire.

EN BREF, L'ÉTUDE ÉMET LES CONSTATATIONS SUIVANTES :

- en dépit des progrès réalisés au cours des 20 dernières années, des améliorations sembleraient encore nécessaires en matière de qualité et d'exploitation des données probantes dans l'action humanitaire internationale ;
- les données probantes sont importantes : l'exploitation de données de bonne qualité permet d'améliorer l'efficacité et la redevabilité de l'action humanitaire et est conforme à l'éthique et aux principes humanitaires ;
- les discussions menées sur les données probantes dans l'action humanitaire peuvent être envisagées dans le cadre d'un plus vaste débat autour de l'établissement et de l'utilisation de données probantes dans les politiques publiques.

1.2 Plan de la présente étude

Cette étude dresse un état des lieux de la qualité des données mises au service de l'action humanitaire à l'heure actuelle. Elle traite principalement des données établies par le système humanitaire international officiel¹ via l'alerte précoce, l'évaluation des besoins, le suivi et le retour d'information, ainsi que la mesure d'impact et les évaluations. Elle examine en outre dans quelle mesure les acteurs du système humanitaire utilisent effectivement les données probantes afin d'aiguiller leurs décisions opérationnelles et la formulation de leurs politiques. Pour évaluer la qualité des données, l'étude se propose d'utiliser six critères principaux :

- Exactitude
- Représentativité
- Pertinence
- Généralisabilité (validité externe) des conclusions
- Attribution
- Clarté autour du contexte et des méthodes

L'étude s'adresse aux responsables des politiques humanitaires, aux décideurs humanitaires (aux niveaux stratégique et opérationnel) ainsi qu'aux personnes concernées par la collecte et l'analyse d'informations dans les organisations humanitaires. Elle est également destinée à être lue par les universitaires et les étudiants en action humanitaire, dans le cadre d'une contribution au débat plus large sur ce thème.

Les questions spécifiques abordées dans la présente étude sont les suivantes :

Section 1: Pourquoi les données probantes sont-elles importantes pour l'action humanitaire ?

Section 2: Qu'entend-on par « données probantes » dans le contexte de l'action humanitaire et comment en juger la qualité ?

Section 3: Les données probantes actuelles satisfont-elles les critères de qualité ?

Section 4: Les données probantes actuelles sont-elles exploitées pour orienter les opérations et l'élaboration des politiques humanitaires ?

Section 5: Comment peut-on améliorer la qualité et l'exploitation des données probantes dans l'action humanitaire ?

L'étude repose sur une analyse bibliographique complétée par des entretiens, et sur le contenu de présentations faites lors de la 28e réunion annuelle d'ALNAP intitulée « Evidence and Knowledge in Humanitarian Action » (Washington DC, mars 2013). Pour plus de détails sur la méthodologie, se reporter à l'annexe 1.

¹ Terme utilisé pour désigner les ONG nationales et internationales, les agences onusiennes, les membres de la famille de la Croix-Rouge et du Croissant-Rouge, et les organismes gouvernementaux hôtes et donateurs impliqués dans le financement, la programmation et la mise en œuvre de l'assistance humanitaire internationale. Le concept de système humanitaire formel repris dans la présente étude est celui qu'utilise ALNAP dans son « État du système humanitaire » : voir (ALNAP, 2012 : 16).

Encadré 1 Définitions

Aux fins de la présente étude, nous adopterons les définitions pratiques suivantes.

Termes clés

Donnée probante : information qui permet de justifier ou de prouver/réfuter le bien-fondé de telle ou telle proposition.

Information : toute donnée susceptible d'informer la compréhension ou une conviction, présentée dans un contexte qui donne du sens à cette donnée. Une information peut être vraie ou fausse. Une information ne devient donnée probante que lorsqu'elle est liée à une proposition spécifique.

Savoir : « conviction ferme et justifiée ». Dans le cadre de la présente étude, le savoir est conçu comme découlant soit de l'observation directe, soit d'une grande quantité de données qui informent la connaissance réelle d'un sujet particulier.

Termes liés à la recherche

Biais : toute forme d'erreur systématique (non aléatoire).

Hypothèse : proposition, étayée par un nombre limité de données probantes, soumise comme explication potentielle d'un phénomène. Une hypothèse est souvent le point de départ d'une enquête future : des données supplémentaires peuvent être recueillies afin de soutenir (ou de réfuter) l'hypothèse.

Qualitative(s) : (recherche, analyse, données) fondée(s) sur des textes plutôt que sur des chiffres. La recherche qualitative tend à avoir trait aux comportements et aux motivations humains.

Quantitative(s) : (recherche, analyse, données) fondée(s) sur des chiffres plutôt que sur des textes. La recherche quantitative repose sur l'analyse (statistique) d'un jeu de données.

Fiabilité : relatif à la cohérence des résultats d'une même expérience lorsque celle-ci est répétée : il s'agit d'une mesure du degré auquel la même expérience produira le même résultat sous les mêmes conditions à différentes occasions.

Méthode scientifique : méthode ou procédure qui caractérise les sciences naturelles depuis le XVII^e siècle ; elle comprend l'observation systématique, la mesure et l'expérimentation ainsi que la formulation, le test et la modification d'hypothèses (d'après l'Oxford English Dictionary).

Validité : terme aux sens divers liés entre eux, tous ayant trait au degré de « bien-fondé » d'un argument expérimental ou de la conclusion d'une expérience. En sciences sociales, et dans la recherche qualitative en général, la validité a tendance à concerner la solidité du lien entre une conclusion ou analyse et les données observées (LeCompte et Goetz, 1982) ou, de manière plus large, la solidité du lien entre une affirmation et le phénomène social auquel elle se rapporte (Hammersley, 1990). En science expérimentale, la validité a plusieurs sens précis. La validité interne est la mesure de la solidité du lien de cause et effet observé au moyen de tests statistiques : le degré auquel un effet particulier est attribuable à une cause particulière. La validité externe, en revanche, examine dans quelle mesure un lien de cause à effet relevé dans une situation particulière est généralisable à d'autres situations.



Qu'entend-on par « données probantes » dans le contexte de l'action humanitaire et comment en juger la qualité ?

2.1 Contexte : termes, concepts et défis

La présente étude repose sur quelques concepts fondamentaux qu'il y a lieu de clarifier dès le départ. Le sens du terme **savoir** a été très débattu par les philosophes à travers les âges, et ce débat est au cœur de la discipline philosophique nommée épistémologie, qui a trait à la nature et à la portée du savoir. Aux fins de notre démarche, nous entendons ce terme au sens de « conviction ferme et justifiée » – autrement dit, une conviction fondée sur une forme quelconque de « fait ». Le savoir diffère de l'information, que nous définissons comme étant toute donnée susceptible d'informer la compréhension ou une conviction, présentée dans un contexte qui lui confère un sens.

Au sein de l'épistémologie, le courant de pensée empirique revêt une pertinence particulière pour notre discussion. Selon les empiristes, le savoir découle d'une expérience reposant sur les sens (« le témoignage de ses propres yeux »). L'empirisme s'oppose à l'idéalisme, à la tradition, à l'autorité : l'idée selon laquelle une chose est correcte parce que la théorie le dit, ou parce que l'on a toujours fait ainsi, ou encore parce que c'est ce que les spécialistes nous disent de faire. C'est cette attitude qui caractérise une grande partie de la pensée scientifique occidentale depuis les quatre siècles derniers, et qui sous-tend la méthode scientifique : formuler des hypothèses, et ensuite tester – et éventuellement modifier – celles-ci au moyen de l'expérimentation et de l'observation. Dans les sciences sociales, l'empirisme est souvent rebaptisé « positivisme », système selon lequel les propositions ne pouvant être testées par l'observation et l'expérience sont dépourvues de sens, à moins qu'elles ne soient vraies par définition ou par déduction logique.

Les démarches scientifiques sociales ne se limitent pas cependant aux méthodes empiriques ou positivistes incorporées aux traditions occidentales dominantes. Depuis le milieu du XXe siècle, le positivisme est remis en question par des points de vue en concurrence les uns avec les autres dans les sciences sociales. Pour l'essentiel, c'est l'existence d'une vérité et d'un savoir « absolu » séparés de l'observateur, et avec elle l'idée d'objectivité stricte, qui a été contestée. Les modèles de cause et effet dits linéaires ont par ailleurs été remis en question. L'accent est mis plutôt sur le relatif et le subjectif, sur les perceptions et les expériences humaines, ainsi que sur des explications plus complexes – et socialement et politiquement informées – des comportements et des résultats. Une telle démarche tend à privilégier des méthodes qualitatives plutôt que des approches quantitatives

strictement axées sur ce qui est mesurable. Les conceptions actuelles de l'anthropologie sociale et culturelle s'inscrivent essentiellement dans ce paradigme et informent bien d'autres domaines de la pensée scientifique sociale et politique.

L'idée de « données probantes » traitée dans la présente étude est étroitement liée à l'empirisme et à la méthode scientifique. Les données probantes sont des informations qui permettent de démontrer le vrai ou le faux d'une hypothèse ou d'une proposition donnée. Souvent, ces propositions portent sur l'existence ou l'absence d'une situation (malnutrition, épidémie, etc.) qui est vérifiable grâce à une observation systématique répétée et menée selon des méthodologies avérées. Dans certains cas toutefois, les propositions concernent le comportement ou les convictions (souvent diverses) de groupes de personnes. Dans ce cas-là, si l'on suit globalement la démarche consistant à élaborer et à tester des propositions, force est de reconnaître qu'il n'existe pas de « réalité » unique et objective et que nous avons plutôt affaire à des idées multiples, voire contradictoires. Il nous faut dès lors étudier la valeur en tant que données probantes de toute une gamme d'informations de types divers, et recourir à diverses méthodologies non expérimentales afin d'examiner l'utilité de ces informations pour étayer la proposition en question.

Si les données probantes humanitaires diffèrent selon leur type, elles varient également en qualité. Des données nutritionnelles recueillies à l'aide d'une méthode valable d'évaluation en la matière constituent de meilleures preuves de l'existence d'une crise alimentaire qu'un rapport empirique faisant état de nombreux enfants émaciés (bien que les deux soient des types de données probantes). Nous devons néanmoins nous garder de supposer que les différences de qualité des données dépendent simplement de la nature (qualitative ou quantitative) de ces données : il se peut par exemple qu'un questionnaire structuré, administré à un échantillon probabiliste de la population, nous en dise moins sur les façons d'améliorer un programme de moyens de subsistance qu'une série d'entretiens semi-structurés bien menés. Selon une publication récente du DFID, *évaluer la solidité d'une donnée* « permet d'éviter de construire une hiérarchie de conceptions et de méthodes de recherche... [et] reconnaît le fait que différentes conceptions sont plus ou moins adaptées à différents contextes et différentes questions de recherche » (DFID, 2013 : 7). Le défi, que nous abordons à la section 2.3, est de recenser des critères pour la qualité des données probantes de divers types d'informations. Dans bien des cas, il nous est possible de produire des données plus solides en utilisant des démarches dites mixtes, en recourant à un large éventail de données reposant sur plusieurs conceptions et méthodes de recherche et triangulant ces méthodes » (ibid : 8).

Enfin, il y a lieu de noter que la qualité des données que nous sommes disposés à admettre diffère aussi en fonction essentiellement des circonstances dans lesquelles nous nous trouvons. Parfois, les données empiriques sont les seules dont nous disposons, et elles suffisent à déclencher une action, même si celle-ci consiste normalement à obtenir des données plus rigoureuses sur lesquelles seront fondées d'autres décisions d'action.



Les données probantes sont des informations qui permettent de démontrer le vrai ou le faux d'une hypothèse ou d'une proposition donnée.



EN BREF :

- Dans la présente étude, nous définissons les données probantes comme étant des informations qui informent une proposition précise.

- Cette conception de la donnée probante s’inspire largement d’une conception empiriste, et essentiellement « scientifique » du savoir.
- Toutefois, en raison de la nature de l’action humanitaire, nous sommes souvent confrontés à des « réalités » subjectives et socialement construites. En conséquence, nous devons nous ouvrir à la nécessité d’admettre comme données probantes de nombreux types d’informations.
- Les données probantes diffèrent en qualité : une telle différence ne dépend néanmoins pas de la nature quantitative ou qualitative des données.
- Le seuil d’exploitation des données probantes dépend des circonstances : dans certaines circonstances, nous devons être disposés à admettre des données de qualité inférieure.

2.2 Des données probantes pour quoi ? Propositions humanitaires et exigences de données probantes

Comme nous l’avons décrit à la section 2.1, toute donnée probante dûment conçue est une donnée au service d’un objet ; plus précisément, il s’agit d’une information ou analyse qui contribue à soutenir une proposition ou une revendication donnée. Puisque le concept de donnée probante est si étroitement lié à l’idée de propositions, il nous faut examiner la nature de ces dernières dans le contexte humanitaire, et les conditions requises afin de démontrer leur caractère vrai ou faux. Nous avançons qu’il est possible d’identifier deux ensembles de propositions liés entre eux qui sont essentiels aux processus de prise de décision des acteurs du système humanitaire officiel :

- Type A : les propositions sur l’existence de besoins humanitaires découlant d’une crise ; et
- Type B : les propositions sur les pratiques efficaces pour répondre à ces besoins.

Chacun de ces deux ensembles comportera diverses propositions spécifiques, et chacune requerra des données probantes (souvent de types différents) pour l’étayer.

Propositions sur l’existence de besoins

Le premier ensemble de propositions (Type A) a trait à l’existence de besoins auxquels doit répondre la communauté internationale. Au sein de cet ensemble, les propositions spécifiques clés sont les suivantes :

- on constate une situation « critique » pour les personnes touchées, ou une situation qui, sans être encore critique, le deviendra si on ne s’emploie pas à la résoudre ;
- cette situation nécessite l’intervention d’acteurs externes.

Ces propositions ont principalement vocation à prouver qu’une intervention humanitaire est requise. Les données probantes qui permettent d’étayer de telles propositions proviennent généralement de l’alerte précoce et d’évaluations, et reposent souvent sur la mesure de certains indicateurs visant à décrire la situation par rapport à certains aspects clés des vies humaines et des moyens de subsistance. Dans bien des cas, les résultats de ces mesures sont comparés aux seuils de crise admis, afin de montrer que les conditions requises de situation de « crise » actuelle ou imminente sont remplies.

La plupart des crises humanitaires – du moins du point de vue des humanitaires professionnels – sont faites de sous-crisis spécifiques à des secteurs : par exemple, des inondations catastrophiques peuvent entraîner des crises de santé publique, de sécurité alimentaire et d’abris. Beaucoup de ces secteurs sont dotés de leurs propres indicateurs, bien que dans certains cas les interventions dépendront moins de déterminer dans quelle mesure les symptômes ont atteint des seuils clairement définis que du jugement du décideur. Le choix du seuil, voire l’absence de seuil, soulève des questions autour de ce qui constitue une crise et quelles données probantes sont requises pour démontrer son existence.

La difficulté à identifier une crise est particulièrement ardue lorsque la proposition porte sur la possibilité d’une urgence à venir plutôt que sur la description de la situation actuelle. Dans ce cas, les données probantes doivent montrer que la situation évolue de telle manière que des conditions d’urgence vont se présenter à l’avenir. Parfois, une telle démonstration est possible en comparant les informations sur la situation actuelle avec des « déclencheurs » prédéterminés d’action précoce. Or, souvent de tels déclencheurs n’existent pas ; en fait, les opinions divergent – et c’est légitime – sur le caractère réaliste de déclencheurs génériques, indépendants de contextes spécifiques.

Dans l’idéal, les propositions de type A nécessitent également des données probantes qui montrent que la crise ne peut pas être résolue efficacement sans une intervention de la communauté internationale. Cependant, en pratique, ce type de preuve n’est pas toujours requis avant le lancement d’une intervention.

Propositions sur l’efficacité de l’intervention

Le second ensemble de propositions (Type B) a trait à la façon dont les acteurs internationaux devraient réagir à une crise, actuelle ou imminente. Au sein de ce groupe, les propositions spécifiques clés sont les suivantes :

- une intervention de type précis permettra (ou a permis) de manière efficace d’empêcher ou d’atténuer les effets de la crise (ou d’une crise similaire) de façons définies ;
- l’intervention proposée est (était) l’action disponible la plus appropriée au contexte, compte tenu de son efficacité probable, du rapport coût-efficacité, de sa pertinence du point de vue de la population locale, etc. (intervention dite de « la meilleure option ») ;
- l’intervention peut être déployée sur la base proposée et satisfait à des standards minimums convenus – autrement dit, elle est faisable.

Ces propositions sont au cœur du choix de réponse qui est fait au moment de décider comment intervenir dans une situation précise. La première compare l’efficacité d’une intervention et l’inaction. La seconde est subordonnée à une démonstration des raisons pour lesquelles l’intervention proposée constitue le meilleur choix parmi les options disponibles ; elle revêt une importance croissante pour les agences humanitaires tandis que celles-ci évaluent les mérites relatifs de diverses méthodes de réponse. Auparavant, beaucoup d’entre elles ont pu recourir à un éventail assez limité de « réponses de défaut » ; or, ces dernières sont désormais remises en questions par les avancées dans les savoirs en matière d’options alternatives. Par exemple, la documentation croissante des enseignements sur les alternatives aux stratégies conventionnelles de lutte contre l’insécurité alimentaire



La difficulté à identifier une crise est particulièrement ardue lorsque la proposition porte sur la possibilité d’une urgence à venir plutôt que sur la description de la situation actuelle.



et de soutien aux moyens d'existence (distribution d'argent au lieu d'aide alimentaire, entre autres) pousse de plus en plus les agences à justifier leurs choix d'intervention (Maxwell et coll., 2012).

La troisième proposition concerne les critères appliqués pour déterminer quelle est la « meilleure » intervention : rapport coût-efficacité, rentabilité, acceptation par la population, capacité à atteindre toutes les personnes dans le besoin, degré de coordination de l'intervention avec d'autres interventions afin de répondre à l'ensemble des besoins de manière cohérente, voire combinaison de ces critères.² Les données probantes liées à ces différents critères requerront des ensembles d'information différents, ainsi que différentes méthodes de collecte et d'analyse.

Des variantes de ces propositions – qui prennent en compte les actions passées – jouent elles aussi un rôle central pour l'évaluation du degré de succès d'une intervention. En général, elles nécessitent des preuves des résultats de l'intervention et présentent le défi souvent complexe de prouver que telle intervention précise a entraîné tels résultats précis.

Données probantes requises afin de traiter les propositions humanitaires

Les propositions de type A et B appellent toutes les deux à deux grands types de données probantes : spécifiques et générales. Les données probantes spécifiques nous informent sur ce qui se passe à un moment particulier dans un lieu précis. Les données probantes générales sont obtenues en étudiant un certain nombre de situations d'urgence similaires afin de pouvoir formuler des avis sensés, par exemple sur la manière dont les famines se développent ou sur le fonctionnement des interventions de santé publique. En examinant ensemble ces deux jeux de preuves, nous pouvons situer un événement dans un contexte plus large et dire par exemple, avec assez de certitude, si les conditions auxquelles nous avons affaire constituent ou non une crise par rapport à des situations observées ailleurs, ou si nous avons devant nous le genre de situation où un type précis d'intervention est souvent efficace.

Dans certains cas, les données probantes générales – concernant le mode de développement des famines ou la meilleure façon de fournir des abris à des réfugiés – sont codifiées en termes de seuils, déclencheurs ou procédures standard, qui reposent tous sur l'idée selon laquelle il est possible d'extrapoler ce qui s'est produit dans d'autres contextes à un contexte particulier. Ces déclencheurs et procédures peuvent accélérer considérablement la vitesse d'intervention et peuvent, surtout, créer un consensus autour de la nécessité d'une intervention et de la forme qu'elle doit prendre. Pour autant, chaque situation d'urgence étant différente, il y a lieu de traiter la plupart des données probantes générales – et les seuils, déclencheurs et procédures qui en découlent – avec une certaine dose de scepticisme, et de s'assurer qu'elles sont adaptées à la situation précise en question.

Il convient également de noter que les données probantes relatives aux deux ensembles de propositions se présentent rarement sous la forme de mesures

² Les critères généralement admis pour déterminer le succès d'une intervention sont ceux proposés dans les critères d'évaluation de l'OCDE-CAD (voir Beck, 2006).

directes. Dans la plupart des cas en effet, nos données prennent la forme d'indicateurs, c'est-à-dire d'informations qui ne décrivent ou ne mesurent pas directement la question concernée, mais qui pointent ou « indiquent » la question. Ainsi les prix, l'état nutritionnel ou les rendements agricoles, s'ils ne fournissent pas une mesure directe de l'insécurité alimentaire, peuvent-ils tous tendre vers celle-ci. En général, des indicateurs sont utilisés, car ils sont (relativement) simples à mesurer et ont un lien raisonnablement fiable et logique avec l'objet que l'on souhaite étudier. Toutefois, ils sont rarement concluants et doivent souvent être utilisés simultanément. La sélection des indicateurs constitue un aspect essentiel des données probantes dans l'action humanitaire, notamment pour les propositions liées au diagnostic (Type A). En effet, lorsque les mauvais indicateurs sont retenus, il est toujours possible d'obtenir des informations parfaitement justes qui n'ont absolument rien à voir avec la proposition concernée.



Lorsque les mauvais indicateurs sont retenus, il est toujours possible d'obtenir des informations parfaitement justes qui n'ont absolument rien à voir avec la proposition concernée.



Avant d'aller plus loin, une mise en garde nous semble nécessaire : les deux jeux de propositions qui sont examinés dans cette étude – et les données probantes auxquelles il est fait appel pour les étayer – résultent des besoins opérationnels des organisations humanitaires internationales. S'ils forment le cœur de cette étude, ils n'englobent ou n'admettent pas forcément tout ce que l'on sait au sujet d'une crise existante. Ce document est rédigé principalement du point de vue des acteurs humanitaires internationaux, mais il ne faut pas oublier que d'autres groupes impliqués dans une situation d'urgence – et notamment les personnes les plus directement touchées – se posent peut-être d'autres questions et il se peut qu'ils exigent d'autres formes de données et de savoirs afin d'y répondre.

EN BREF :

- La plupart des données probantes recueillies par les organisations humanitaires internationales ont pour but d'aborder des propositions sur l'existence de besoins ; ou des propositions sur la meilleure façon de répondre à ces besoins.
- Les données probantes relatives à ces propositions se présentent sous diverses formes et ont trait à de nombreuses questions différentes.
- Les décisions humanitaires sont souvent prises en comparant les données issues d'une situation précise avec des données « générales » portant sur la manière dont des situations similaires ont évolué ou ont été traitées auparavant ; la prudence s'impose toutefois : les informations « générales » qui résultent d'interventions antérieures ne sont pas toujours pertinentes.
- Un grand nombre d'informations exploitées par les humanitaires prennent la forme d'indicateurs, et il est impératif de choisir ces derniers avec soin pour s'assurer qu'ils ont un lien solide et logique avec la situation qui est évaluée.

2.3 Critères d'évaluation de la qualité des données probantes humanitaires

Lorsque nous tentons de justifier les propositions présentées dans la section 2.2, nous nous heurtons à plusieurs problèmes. D'une part, nous devons être sûrs que les propositions sont vraies, à partir des meilleures données probantes disponibles. Or, comment savoir si nous pouvons nous fier à ces données ? D'autre part, comment juger si les données probantes étayent effectivement la proposition

en question ? Dans cette section, nous examinons quelques-unes des questions importantes et proposons six critères de « données solides ».

Dans les contextes qui nous occupent, les données probantes se présentent à la fois sous forme de données, quantitatives ou qualitatives, directes ou indirectes, et sous forme de résultats d'analyse de ces données, tels que les conclusions d'évaluations, d'appréciations ou d'études. Lorsque nous essayons de déterminer si des informations peuvent servir de données probantes – c'est-à-dire si elles permettent de donner plus de crédibilité à une proposition ou de la réfuter –, nous devons garder à l'esprit que ce n'est pas seulement la qualité des données qui est importante, mais aussi la qualité de l'analyse. Il est en effet tout à fait possible que des données valables soient mal analysées et que les mauvaises conclusions en soient tirées. Avant d'affirmer que des informations comptent en tant que données probantes, il y a lieu de mesurer les données et les méthodes utilisées pour les analyser par rapport des critères de qualité spécifiques.

Comme nous l'observons plus haut, on peut décrire des données probantes de différents types en termes de plusieurs attributs clés : véricité, crédibilité, exactitude, fiabilité et validité. Hélas, ces termes sont utilisés de manières différentes entre divers disciplines et secteurs, ce qui rend difficile toute généralisation de leur sens. Pour compliquer davantage le problème, certains de ces termes sont employés pour décrire à la fois la qualité des données et les méthodes de collecte et d'analyse de ces données.

Nous nous sommes efforcés ici de résumer les six principaux critères susceptibles de servir à évaluer la qualité des données probantes dans des contextes humanitaires.

EXACTITUDE



- i. **Exactitude** : elle permet de déterminer si les données probantes correspondent bien à une situation réelle, et si elles rendent compte fidèlement du phénomène qui est mesuré. Des mesures anthropométriques correctement effectuées et des déclarations d'informateurs clés qui rendent vraiment compte de ce que croient ces derniers répondent toutes les deux aux critères d'exactitude. L'exactitude peut être compromise par l'informateur ou par l'observateur/chercheur, par une falsification délibérée, un biais inconscient, un malentendu ou un usage incorrect des instruments de mesure.

REPRÉSENTATIVITÉ



- ii. **Représentativité** des données probantes : c'est le degré d'exactitude de représentation par les données probantes de l'état du groupe plus large visé. Par exemple, des informations qui remplissent les critères d'exactitude pour refléter la situation dans un village ou les points de vue d'un groupe de population ne représentent pas forcément tous les villages ou tous les groupes de la région concernée. C'est la question de savoir dans quelle mesure un échantillon représente l'ensemble de la population

PERTINENCE



- iii. **Pertinence** : il s'agit du degré de solidité du lien entre une information et la proposition qu'elle est censée prouver ou réfuter. Une information peut en effet être exacte et représentative, mais non pertinente par rapport à l'argument ou à la proposition avancée. La question de la pertinence est souvent liée aux indicateurs, qui ne sont pas des mesures directes d'un état, mais qui mesurent un phénomène tendant vers cet état, ou indiquant sa présence. Les mesures anthropométriques fournissent des indications fiables, par exemple, de l'état

nutritionnel : il existe un lien solide entre l'indicateur et l'état de la situation, aussi ces mesures sont-elles pertinentes en tant que données probantes pour prouver la proposition « il existe un état de malnutrition dans cette région ». Les mêmes mesures sont cependant moins efficaces en tant qu'indicateurs de pénurie alimentaire, car le lien entre la circonférence du bras et le manque de nourriture est moins fort (la malnutrition pourrait être causée par une maladie ou de mauvaises pratiques alimentaires). En conséquence, les tests du MUAC sont moins pertinents en tant que données visant à prouver la proposition « une aide alimentaire est requise ». De manière plus générale, le test de *pertinence* devrait être appliqué à toute information fournie comme preuve à l'appui d'un argument ou d'une proposition particulière.

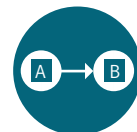
iv. Généralisabilité des conclusions : elle permet de déterminer dans quelle mesure les données probantes issues d'une situation précise peuvent être généralisées au-delà de cette intervention à d'autres situations (et en particulier à des situations dans d'autres pays et à d'autres moments), et ainsi servir d'indications plus générales de la manière dont une situation évoluera, ou du type d'intervention le mieux adapté. La question de généralisabilité est particulièrement importante pour les dirigeants, qui tentent d'élaborer des politiques en appliquant globalement les données probantes obtenues dans des contextes précis. Elle est en outre au cœur du débat sur les indicateurs et les seuils communs mondiaux pour les catastrophes.

GÉNÉRALISABILITÉ



v. Attribution de l'analyse : il s'agit de déterminer si l'analyse démontre un lien de causalité clair et sans équivoque entre deux états ou événements. Ce concept est fondamental pour le caractère probant de l'évaluation : lorsque des évaluateurs relèvent certains changements de situation qui sont survenus après une intervention, comment peuvent-ils savoir avec certitude que ces changements ont été le résultat de cette intervention précise, et non d'un autre facteur ?

ATTRIBUTION



vi. Clarté autour du contexte et des méthodes : dans quelle mesure il est expliqué clairement pourquoi, comment et pour qui les données probantes ont été recueillies. Une donnée probante, nous l'avons vu, est une information qui est liée à une proposition spécifique. À ce titre, une information ne constitue une preuve que dans le contexte d'une question précise, posée par un organisme ou un groupe particulier. Un observateur ne peut évaluer le caractère probant d'une information que s'il sait à quelle proposition elle se rapporte, qui souhaitait prouver cette proposition et comment l'information a été recueillie. Faute d'informations sur le contexte, il est impossible d'établir si les données probantes sont *pertinentes* ou *généralisables*. De même, une information peut être admise comme donnée probante uniquement si les méthodes employées pour sa collecte et son analyse, et les restrictions éventuelles liées au processus, sont précisées. Ce n'est qu'à partir de ces informations que l'utilisateur peut déterminer l'*exactitude*, la *représentativité*, la *pertinence*, la *généralisabilité* et l'*attribution* des « données probantes ».

CLARTÉ AUTOUR DU
CONTEXTE ET DES
MÉTHODES

Cette classification des critères ne fait en aucun cas autorité, et il existe plusieurs autres démarches permettant de mettre à l'épreuve la qualité des données probantes. Dans une étude de 2005 sur l'exploitation des données pour l'élaboration des politiques, Louise Shaxson, s'appuyant sur les travaux de Spencer et coll. (2003), propose cinq composantes qui, ensemble, définissent la « robustesse » ou solidité des données probantes sur le plan politique : crédibilité,

fiabilité, objectivité (absence de biais), enracinement³, et *généralisabilité* (Shaxson, 2005). De son côté, le groupe *British Overseas NGOs for Development* (BOND) propose cinq principes pour évaluer la solidité des données probantes présentées par les ONG (secteurs du développement et de l'humanitaire) : degré d'inclusion dans les données probantes des points de vue des populations vivant dans la pauvreté ; obtention des données par des méthodes qui sont justifiables au regard de la nature de l'évaluation ; degré de triangulation des données ; capacité des données probantes à examiner comment les changements se produisent ; enfin, capacité des données à révéler les détails des sources de données et des méthodes employées, les résultats obtenus, et les limites éventuelles des données ou des conclusions.

Proposant un cadre unifié pour les méthodologies socioscientifiques, John Gerring avance certains critères applicables à tous les arguments de sciences sociales (propositions) : véracité, précision, généralité, cohérence, commensurabilité et pertinence. Toutefois, comme il en fait l'observation : « Malheureusement, le vocabulaire associé à la question de la méthodologie abonde d'ambiguïtés. La terminologie clé... revêt des sens divers dans différents contextes et traditions de recherche » (Gerring, 2011 : 16). Il ne faut donc pas s'étonner si une liste unique de critères communs est difficile à dresser. Nous espérons toutefois que les six critères proposés ici, qui reflètent le consensus auquel sont parvenus les professionnels de l'action humanitaire et les universitaires lors de la 28e réunion annuelle d'ALNAP⁴, fourniront un instrument utile aux acteurs humanitaires lorsqu'ils examinent la solidité des données probantes sur lesquelles ils s'appuient.

EN BREF :

- S'il existe différents ensembles de critères pour juger la qualité des données probantes, ils portent souvent sur un type précis de données (qualitative/ quantitative) et ne sont donc pas généralement applicables à toutes les données probantes humanitaires.
- Le présent rapport mise sur six critères pour évaluer la qualité des données probantes obtenues et exploitées dans l'action humanitaire : *exactitude, représentativité, pertinence, généralisabilité, attribution et clarté autour du contexte et des méthodes.*

3 L'enracinement consiste à savoir si la question sur laquelle portent les données probantes représente véritablement l'intégralité du problème concerné ou si d'autres aspects pourraient et devraient être étudiés. Cette idée est liée à celle de *clarté autour du contexte et des méthodes* présentée ci-dessus.

4 Une liste initiale, comprenant en gros les cinq premiers critères, a été présentée à la réunion et débattue en plénière. Si aucune objection aux critères eux-mêmes n'a été soulevée, on estimait en général qu'ils ne tenaient pas compte des destinataires des données probantes, et du but de la collecte de celles-ci. Le sixième critère visait à répondre à cette préoccupation. Les délégués de l'assemblée ont par ailleurs proposé que le critère de délai d'obtention soit adopté. Ce critère n'a pas été retenu dans la présente étude, car, selon les auteurs, il semble davantage lié à l'utilité de la donnée probante plutôt qu'à la qualité intrinsèque de celle-ci.

3

Les données probantes humanitaires répondent-elles à l'heure actuelle aux critères de qualité ?

3.1 Les enjeux de la production de données probantes dans les situations humanitaires

Les sérieux défis à relever pour produire des données valables dans le secteur humanitaire sont nombreux. Beaucoup découlent du contexte au sein duquel les acteurs humanitaires opèrent. Les acteurs internationaux se trouvent souvent impliqués dans des interventions où les institutions de l'État sont plutôt faibles. En conséquence, les données secondaires sont souvent très limitées dans les zones d'intervention des humanitaires : même des données démographiques de base y sont difficiles à obtenir (encadré 2).



Malheureusement, les circonstances qui rendent la collecte de données si importante sont précisément celles qui rendent cette tâche extrêmement difficile.



Lorsque des données secondaires utiles sont disponibles, la nature des situations d'urgence humanitaire – qui se caractérisent souvent par des niveaux élevés de mortalité et d'importants déplacements de population – les rend très vite obsolètes. Les chiffres démographiques, les taux de malnutrition, les données sur les abris : autant de données qui sont susceptibles d'évoluer très rapidement, rendant nécessaire de recueillir l'information en continu au fil de la progression de la situation.

Malheureusement, les circonstances qui rendent la collecte de données si importante sont précisément celles qui rendent cette tâche extrêmement difficile. Beaucoup d'interventions humanitaires ont lieu dans des situations où l'accès physique est sérieusement restreint, ce qui limite les possibilités de recueil des données. En même temps, la distribution de grandes quantités de biens et services dans des contextes où les besoins sont énormes a, inévitablement, des implications politiques, notamment dans les situations d'urgence dites complexes où des fractures entre groupes politiques et entre factions existent déjà dans une région donnée. Dans de telles circonstances, les humanitaires ne peuvent pas supposer que les informations – et en particulier celles recueillies auprès d'informateurs clés – sont objectives et exactes. Et ils ne peuvent pas non plus automatiquement s'estimer eux-mêmes impartiaux.

ENCADRÉ 2 La question des chiffres

Certaines des données chiffrées de base dont dépendent les propositions humanitaires sont notoirement arbitraires. En particulier, les estimations de population et les chiffres fournis pour les « nombres de personnes affectées » ou les « nombres de bénéficiaires atteints » ont tendance à souffrir d'un degré élevé d'incertitude et d'un manque de clarté des définitions. Ce flou risque alors d'ébranler la crédibilité des propositions qui font référence à ces chiffres.

La racine de ces problèmes est fréquemment dans l'incertitude qui entoure les chiffres sur la population de base (dénominateur). Normalement, c'est le recensement de la population qui constitue la meilleure source de données démographiques de base. Or, il peut arriver qu'il n'y ait pas de recensement disponible à l'échelon local. De plus, lorsque le recensement est utilisé comme instrument politique, les chiffres qu'il fournit ne sont pas fiables. Même lorsque ce n'est pas le cas, les chiffres issus d'un recensement sont souvent dépassés, et bien qu'il existe des méthodes traditionnelles pour créer des projections de population, celles-ci misent sur des hypothèses critiques difficiles à émettre avec une quelconque certitude.

Faute de données issues d'un recensement, les humanitaires ont eu recours à diverses approches : comptage rapide – soit sur le terrain, soit à l'aide d'imagerie satellitaire – des habitations et des bâtiments ; dénombrement des effectifs ; établissement de listes et de profils ; surveillance des mouvements ; étude randomisée sur les ménages. De telles méthodes demandent toutefois du temps et varient en *exactitude*. Ce qui est frustrant, c'est qu'elles ont tendance à être plus efficaces lorsque des données de recensement sont disponibles.

Conséquence fâcheuse de cette carence de données de base : la diversité au sein de la société est souvent masquée. En particulier, les différences entre les sexes et les âges, élément important de la plupart des recensements, sont rarement recueillies parmi les groupes sociaux évalués et la ventilation des données selon ces critères demeure l'exception plutôt que la norme.

Sources : Demographic Assessment Techniques in Complex Humanitarian Emergencies: Summary of a Workshop (2002) Committee on Population, National Academies Press, US ; 'Sex and Age Matter', Tufts University (2011) ; État du système humanitaire, ALNAP (2012) ; Guidelines on Data Issues in Humanitarian Crisis Situations, UNFPA (2010) ; Technical Brief: Estimation of Affected Population Figures (ACAPS, 2012).

Les opérations humanitaires sont par ailleurs marquées par les différences de pouvoir qui existent entre les tiers (qui contrôlent les ressources et comprennent le système humanitaire) et les personnes affectées par la situation d'urgence. Si ces dernières risquent d'avoir des besoins énormes, elles ont tendance à n'avoir qu'une compréhension limitée de la façon dont le système fonctionne et n'ont que très peu de temps pour remplir des questionnaires ou participer à des entretiens. Non seulement cela crée des difficultés pratiques considérables pour la collecte des données primaires, mais cela entraîne des difficultés d'ordre éthique préoccupantes : dans quelle mesure est-il légitime en effet de demander à la population de fournir des informations dont elle risque de ne pas bénéficier elle-même directement ? Dans quelle mesure peut-on s'attendre à ce que la population affectée donne son consentement informé à l'utilisation et à la publication d'informations dans un système qu'elle ne comprend pas, et dont elle espère qu'il lui apportera de l'aide matérielle ? Comme l'affirme Bradt, « la collecte de données et les interventions humanitaires qui s'ensuivent constituent des procédures invasives aux conséquences imprévues. Les bonnes intentions n'excusent pas les mauvais résultats » (Bradt, 2009 : 15).

Autre défi éthique : la production de données probantes nécessite des fonds (et du temps) qui, dans la plupart des cas, proviennent de budgets qui pourraient être affectés au soutien direct des populations touchées. Cela crée une tension difficile entre le besoin de données rigoureuses et le besoin d'intervenir.

Enfin, les opérations humanitaires se caractérisent souvent par le besoin d'obtenir des informations dans des délais très brefs. Il convient toutefois de ne pas exagérer ce point : les États où la plupart des dépenses humanitaires ont été effectuées au cours des cinq dernières années sont tous en proie à des conflits chroniques (ALNAP, 2012), et une part importante de l'action humanitaire menée dans ces pays prend la forme de programmes de longue haleine au sein desquels le temps pour collecter et analyser les données ne manque pas (Macrae, 2013). Toutefois, dans les situations d'urgence à survenue rapide, telles que le tremblement de terre en Haïti en 2010 ou le typhon Yolanda (Haiyan) en 2013, le temps est compté, ce qui limite la capacité des acteurs humanitaires à recueillir assez de données probantes valables pour soutenir une action efficace dans les premières phases d'une intervention.



Les opérations humanitaires en Afghanistan et au Zimbabwe diffèrent considérablement tant par leur contexte que par leur nature. Dans de telles circonstances, établir les preuves de ce qui est efficace s'avère extrêmement difficile.



Toutes ces contraintes découlent principalement du contexte de l'action humanitaire à l'échelon d'un projet ou d'un programme individuel. Cependant, des problèmes se posent aussi à l'échelon mondial ou systémique – notamment parce que les opérations humanitaires en Afghanistan et au Zimbabwe diffèrent considérablement tant par leur contexte que par leur nature. Dans de telles circonstances, établir les preuves de ce qui est efficace s'avère extrêmement difficile, dans un secteur où les variables socioculturelles sont vitales pour le succès de tant d'opérations.

Dans le reste de cette section, nous examinerons comment les organisations affrontent ces défis, en observant comment les données probantes et les savoirs sont produits concrètement dans le secteur humanitaire. Nous nous pencherons également sur d'autres questions du même ordre concernant la qualité de ces données probantes. Nous accordons une attention particulière aux pratiques actuelles dans le secteur, dans la mesure où elles sont liées aux éléments « standard » de la programmation humanitaire : les processus de production de données des systèmes d'alerte précoce et de surveillance, les procédures d'évaluations des besoins et de suivi, et les évaluations de projet ou de programme. Nous examinerons la nature des données produites par chacune, notamment par rapport aux critères de « qualité » présentés à la section 2 plus haut.

EN BREF :

- En général, il existe un besoin urgent d'informations dans les situations d'urgence et les crises.
- Toutefois, dans de nombreux contextes où l'aide internationale est requise, seules des données de base limitées sont disponibles.
- La collecte d'informations dans les situations d'urgence peut être extrêmement difficile en raison de contraintes d'accès, de l'utilisation de l'information comme instrument politique, et des importantes considérations éthiques à prendre en compte.

3.2 Données probantes émanant des systèmes d'alerte précoce

Les systèmes d'alerte précoce ont été décrits comme des « combinaisons d'instruments et de procédures inscrites dans des structures institutionnelles... [et] composées de quatre éléments : connaissances du risque, service de surveillance et d'alerte technique, diffusion d'avertissements utiles à destination des populations

exposées au risque, et sensibilisation et préparation du public à agir ».⁵ Tous les processus d'alerte précoce ne font pas partie d'un tel système cloisonné. Certains, les processus d'alerte précoce au risque de famine notamment, fournissent des services indépendants de collecte et d'analyse des données sans prescrire une action à mener. D'autres sont moins formels et plus ancrés dans la communauté.

Les systèmes d'alerte précoce combinent les nouvelles données sur une situation en train de se développer avec des connaissances historiques afin de prédire une issue probable pour une région ou un délai donné. Bien que les données concernées soient liées à un contexte précis, l'analyse s'appuie sur une connaissance à la fois du contexte (par exemple, les types de moyens de subsistance) et des événements antérieurs survenus dans ce contexte ou dans d'autres afin d'émettre des prédictions spécifiques à la situation. Cette analyse repose souvent sur des tendances : autrement dit, sa solidité dépend de sa capacité à monter un argument convaincant en faveur d'une tendance émergente qui – si elle n'est pas maîtrisée – aura des conséquences catastrophiques (proposition de Type A ci-dessus).

L'alerte précoce au risque de famine est l'exemple le mieux connu de ce type. L'argument à défendre est normalement complexe, mêlant données de types divers, tels que les cycles de précipitations, les rendements des récoltes, les prix des denrées alimentaires, les termes de l'échange, les revenus des ménages et les niveaux de malnutrition. D'autres arguments sont plus simples, par exemple l'alerte précoce qui suit la trajectoire vraisemblable d'un cyclone qui approche ou la survenue probable d'une inondation, voire d'un tremblement de terre imminent ou d'une éruption volcanique, quoique ces derniers soient plus difficiles à prédire.⁶ L'alerte précoce permet de gagner du temps et de prendre des mesures préventives, préparatoires ou d'évitement.

L'alerte précoce est l'un des domaines de l'action humanitaire qui a réalisé le plus de progrès au cours des 20 dernières années. Des systèmes d'alerte précoce contre la famine en Afrique, tels que FEWS Net et le Cadre intégré de classification de la FAO (IPC), aux systèmes de surveillance des cyclones en Asie et dans les Caraïbes, les avancées technologiques alliées à des systèmes nationaux et transfrontaliers coordonnés et à une mobilisation communautaire efficace ont contribué à réduire la vulnérabilité. Associés à d'autres éléments de préparation (par exemple, abris contre les inondations au Bangladesh, distribution d'aide alimentaire dans la Corne d'Afrique, mécanismes d'intervention communautaire en Inde et en Amérique centrale), ces systèmes ont permis de sauver de nombreuses vies humaines.

Les données probantes qui émanent de la plupart des systèmes d'alerte précoce établis semblent en général être bien notées pour leur *exactitude* lorsqu'elles sont évaluées par rapport aux critères de solidité des données probantes. De nombreux systèmes sont cependant moins performants par rapport à la question de *pertinence* – le lien entre les données qui ont été recueillies et la proposition selon laquelle une crise s'annonce qui aura un impact majeur sur les vies humaines et les moyens de subsistance.

5 Source AlertNet : <http://www.trust.org/alertnet/news/early-warning-of-disasters-facts-and-figures>

6 UNISDR, 2006.



Une telle complexité ne facilite pas non plus une description claire et convaincante de la situation destinée aux décideurs.



Dans les crises à survenue rapide où, par exemple, des cyclones et des inondations peuvent être étroitement surveillés, il est de plus en plus possible de déterminer avec quasi-certitude si les circonstances de ces événements ont des chances d'avoir un impact humain catastrophique et quelles régions seront touchées. En revanche, dans le cas des catastrophes à évolution lente, telles les crises de sécurité alimentaire liées à la sécheresse, de telles prévisions sont plus difficiles. En effet, une multitude de facteurs se conjuguent dans ces crises pour en déterminer les répercussions sur les populations et leurs moyens d'existence, ce qui rend l'importance à accorder aux indicateurs et leur choix extrêmement difficile : les rendements des récoltes ou les cours des céréales sont-ils des variables explicatives de l'insécurité alimentaire future ? Comment les deux sont-ils liés ? Cette interaction complexe de facteurs a pour autre conséquence que des indicateurs négatifs n'entraînent pas toujours de mauvaises situations, et des facteurs imprévus entrent souvent en jeu. Une telle complexité ne facilite pas non plus une description claire et convaincante de la situation destinée aux décideurs.

L'intervention tardive dans la crise d'insécurité alimentaire et de subsistance dans la Corne d'Afrique en 2010/11 a souligné ces difficultés et mis en lumière les divers systèmes d'alerte précoce en opération dans la région. Dans l'ensemble, les observateurs ont conclu que les systèmes « fournissaient des informations exactes et rapides qui permettaient aux autorités de planifier et d'intervenir » (Hillier et Dempsey, 2012 : 14. Voir aussi Majid, 2011 ; Slim, 2012). Un rapport conjoint d'Oxfam GB/Save the Children fait observer que « les rapports de FEWS Net et du groupe FSNWG ont été évalués 'très bien' à 'excellent' sur le plan de l'exactitude dans leur prédiction de la gravité et de la survenue de la crise »⁷ (Hillier et Dempsey, 2012 : 14. Voir aussi Majid, 2011 ; Slim, 2012 : 14).

Toutefois, même dans ces systèmes efficaces, le caractère parcellaire des données a amené certaines personnes à remettre en question l'*exactitude* et la *représentativité* des rapports sur l'alerte précoce (Levine et coll., 2011). Cette difficulté est aggravée par le fait que les méthodes d'analyse de cette multitude d'informations complexes n'étaient pas transparentes – la *clarté autour du contexte et des méthodes* faisait défaut (ibid). Les auteurs du rapport *Échec du système : réévaluation du problème de rapidité de l'intervention dans les crises de la Corne d'Afrique (System Failure: Revisiting the problem of timely response to crises in the Horn of Africa)* reconnaissent que la seule façon de résoudre ce problème d'analyse et d'interprétation serait de trouver « le parfait indicateur qui ne peut pas être manipulé et qui nous dira – dans chaque district – quelle intervention est requise » (Levine et coll., 2011 : 11). Or, la solution à un ensemble de problèmes portant sur les données probantes créerait un autre problème, car aucun indicateur unique ne pourrait cerner avec exactitude l'insécurité des moyens de subsistance dans tous les contextes (par rapport à la typologie présentée plus haut, aucun indicateur unique ne serait *pertinent* dans tous les cas : c'est un problème de *généralisabilité*).

Paradoxalement, les débats – visant à améliorer la qualité des données – sur les méthodologies et indicateurs les mieux adaptés à l'alerte précoce semblent avoir rendu les décideurs non pas moins, mais plus sceptiques quant à l'intérêt des systèmes d'alerte précoce. Nous reviendrons sur le thème de l'usage qu'ils font des

⁷ Dans une certaine mesure, ceci est peut-être dû au fait que l'alerte précoce a trait à l'incertitude : à des situations qui pourraient survenir à l'avenir, plutôt qu'à des situations évaluables immédiatement. Cette incertitude amène les décideurs à exiger un niveau de preuve, et donc de certitude, plus élevé avant d'être disposés à agir.

données probantes – et notamment des données en provenance des systèmes d'alerte précoce – dans la section 4.

EN BREF :

- L'alerte précoce est un domaine qui a fait l'objet d'investissements et d'améliorations considérables ces dernières années.
- Dans ce domaine, la principale difficulté liée aux données probantes est de sélectionner des indicateurs pertinents capables de prédire les événements de manière fiable, en particulier dans les cas où l'alerte précoce concerne des crises (famines, etc.) qui découlent de nombreux facteurs en interaction dans le temps.
- Pour être efficaces, les systèmes d'alerte précoce doivent clairement définir leurs méthodes, notamment d'analyse et d'évaluation de nombreux types de données probantes.

3.3 Données probantes issues de l'évaluation des besoins

On entend ici par « évaluation des besoins » toute un éventail d'activités qui visent à déterminer si une assistance extérieure est nécessaire et, dans l'affirmative, à définir le type, l'échelle et le calendrier de celle-ci. Dans le cadre de la présente étude, la plupart des évaluations des besoins visent donc à fournir des données probantes liées aux propositions formulées sur l'existence d'une crise (propositions de type A) et, dans une moindre mesure, sur la meilleure façon de gérer une crise (type B).

Les évaluations des besoins souffrent des contraintes génériques liées à la collecte des données probantes décrites à la section 3.1 : carence de données de base, difficulté d'accès, manque de temps. Cependant, elles se heurtent aussi aux difficultés spécifiques qui découlent de la nature même de la tâche. Surtout, les situations d'urgence ont tendance à produire une gamme étendue de types de besoins disparates. Nous pouvons prédire certains d'entre eux – en particulier ceux liés à la santé et à la nutrition – et en donner une bonne représentation grâce à des mesures d'indicateurs objectivement vérifiables faisant appel à des méthodes et des protocoles standard. Toutefois, certains besoins, telle l'insécurité alimentaire, constituent des catégories conceptuelles, lesquelles sont mesurées par diverses combinaisons d'indicateurs. Comme nous l'avons fait observer dans la section sur l'alerte précoce, ces indicateurs divergent souvent d'un lieu à l'autre et, en conséquence, très peu de gens s'accordent sur lesquels utiliser et sur le poids qu'il faut leur donner. D'autres besoins comportent un fort élément de subjectivité : les sentiments de menace ou d'insécurité, par exemple, sont signe de besoins qui méritent de l'attention, mais qui sont difficiles à représenter par des méthodes de mesure quantitatives (Mahmood et coll., 2010).

En conséquence, les évaluations de besoins requièrent – idéalement – des individus ou des équipes compétents dans diverses disciplines, capables de recueillir, analyser et rapprocher différents types de données. En pratique, à cause de ce besoin de synthétiser les données probantes et de combler les lacunes éventuelles en matière de données, beaucoup d'évaluations des besoins dépendent lourdement des connaissances contextuelles des individus impliqués dans l'intervention (voir Poole et Primrose, 2010 ; Assessment Capacities Project (ACAPS), 2012 ; CICR et FICR, 2008).

La connaissance du contexte est également importante pour déterminer l'ampleur requise de l'évaluation. Dans certains cas, lorsque des catastrophes de même nature se répètent et ont des conséquences similaires, cette connaissance du terrain peut permettre aux humanitaires de minimiser l'évaluation, voire de s'en dispenser et de passer directement à la programmation. Elle peut aussi les aider à identifier les méthodes les mieux adaptées et à approcher l'évaluation de la meilleure façon en un lieu donné. La disponibilité de l'information et du personnel compétent et la possibilité d'accéder à la zone d'urgence diffèrent toutes d'une situation à une autre. En conséquence, « il n'y a pas de taille unique... il n'existe pas de programme [d'évaluation] unique standard qui puisse être déployé à chaque catastrophe. Il s'agit plutôt d'adapter les pratiques de référence en matière de méthodes efficaces d'évaluation aux besoins d'information propres à chaque crise » (Assessment Capacities Project (ACAPS), 2012 : 6).

Ceci présente néanmoins un risque : faute de méthode d'évaluation unique et standardisée, chaque agence créera dans chaque situation d'urgence sa propre méthodologie d'évaluation applicable à sa propre évaluation séparée. Un certain degré de duplication dans l'évaluation – surtout lorsque cette duplication est planifiée et coordonnée – est sans doute souhaitable : les évaluations multiples peuvent en effet accroître la qualité probante de l'information en permettant de recouper les résultats. Pour autant, l'expérience passée tend à montrer que « la plupart des agences avaient leurs propres formulaires d'enquête non standardisés... qui produisaient souvent des résultats contradictoires ou répétitifs » (Darcy et coll., 2013 : 24), ce qui ne constitue bien évidemment pas une utilisation efficace de rares ressources. Diverses agences recueillent des informations similaires, tout en ne parvenant pas à combler des lacunes évidentes⁸ ; à maintes reprises, les habitants des communautés touchées par des catastrophes passent un temps précieux à répondre aux mêmes questions. De plus, les rapports qui en découlent ne sont exploités que par l'agence qui les a préparés, et ce pour plusieurs raisons possibles : les rapports ne sont pas partagés (Mills, 2005) ; ils ne sont pas partagés en temps utile (Darcy et Garfield, 2011) ; ils ne sont pas pertinents pour les autres parties prenantes (Bradt, 2009) ou ne se présentent pas dans un format utile aux autres acteurs (Poole et Primrose, 2010).

“

Sans doute faut-il rechercher la réponse dans une coordination accrue entre les évaluations et dans la création et l'adoption de normes et de méthodologies communes pour la collecte des données.

”

Sans doute faut-il rechercher la réponse dans une coordination accrue entre les évaluations et dans la création et l'adoption de normes et de méthodologies communes pour la collecte des données (DFID 2012 : 31, Poole et Primrose, 2010 : 1) qui, tout en étant adaptables à des contextes précis, utilisent globalement un langage, des indicateurs et des démarches similaires. Heureusement, de plus gros efforts ont été consentis ces dernières années pour accroître la coordination et élaborer des normes générales, avec le soutien d'acteurs tels que la FICR/CICR, le Groupe de travail sur l'évaluation des besoins (NATF) de la CPI, et l'Assessment Capacities Project (ACAPS).

Sur le plan de la coordination, la tendance générale serait à une acceptation des évaluations communes ou coordonnées qui, comme le font observer Garfield et ses collaborateurs, ont le potentiel d'améliorer l'efficacité et de fournir une base plus solide à une planification et une action coordonnées (avec toutefois un impact potentiel au niveau de la rapidité et des dépenses) (Garfield et coll., 2011). Selon une

⁸ Par exemple, le recueil de données ventilées selon l'âge et le sexe, voir Mazurana et coll., 2011.

estimation citée dans l'État du système humanitaire, une évaluation des besoins menée conjointement par les agences qui participaient au projet de renforcement des capacités d'urgence (ECB) dans l'Est de Java et à Sumatra a permis d'intervenir sur une zone géographique plus étendue et d'accroître le nombre de secteurs couverts (Wilson et coll., 2010).

Entre-temps, un bilan des orientations fournies par divers « acteurs clés » spécialisés dans l'évaluation et des rapports d'évaluation publiés récemment⁹ semblent indiquer un consensus méthodologique croissant autour des approches, et notamment de celles visant les évaluations précoces multisectorielles (effectuées au cours des trois ou quatre premières semaines après une catastrophe). Le consensus qui se dessine semble être en faveur d'évaluations fondées sur un examen des données secondaires – si possible – auxquelles s'ajoutent les informations recueillies à l'aide de questionnaires auprès d'informateurs clés et de groupes cibles de la population affectée. Dans la plupart des cas, les communautés au sein desquelles des entretiens doivent être menés sont sélectionnées selon un échantillonnage au jugé plutôt que selon un échantillon aléatoire. Si les orientations du CICR/FICR proposent d'ores et déjà un format de questionnaire (CICR et FICR, 2008), les recommandations de l'Évaluation multisectorielle rapide (abrégé en MIRA) du groupe de travail sur l'évaluation des besoins restent provisoires (CPI, 2012a) et sont en cours d'amélioration (Joyce Luma, communication personnelle, avril 2013). L'Assessment Capacities Project (ACAPS), lui, ne fournit pas de format de questionnaire, mais pourrait le faire à l'avenir (Richard Garfield, communication personnelle, avril 2013). Dans tous les cas, ces formats ne sont pas prescriptifs, mais visent à tenir lieu de « points de départ auxquels les équipes [devraient] ajouter d'autres questions » (Joyce Luma, communication personnelle, avril 2013).

Si nous envisageons l'état actuel de l'évaluation des besoins humanitaires à la lumière des critères de qualité des données probantes proposés à la section 2.3, il est peut-être inévitable que les évaluations des besoins – qui formeront la base des distributions de quantités limitées de ressources rares et indispensables – se heurtent à des problèmes d'*exactitude*. Celles-ci encouragent évidemment les gens à exagérer la gravité d'une situation donnée afin d'obtenir un niveau accru de soutien. En même temps, il se peut que les gouvernements sous-estiment l'étendue des besoins pour nier qu'une crise a lieu.

En général, les orientations en matière d'évaluation des besoins permettent de penser que la meilleure façon de relever ce défi d'*exactitude* est de trianguler les informations fournies par les informateurs clés ou les groupes cibles avec les résultats d'autres groupes ou informateurs ainsi qu'avec des sources secondaires, voire, dans certains cas, avec certaines « constantes » telles que l'apport calorifique requis (voir Seaman et coll., 2000). Une évaluation de la situation à Aleppo (Syrie) nous offre un exemple de triangulation rigoureuse : les agents recenseurs

9 Les rapports examinés étaient les suivants : The Philippines Multi-Cluster Needs assessment for Tropical storm Washi; Multi-Sector Initial Rapid Assessment, Pakistan Floods, 2012; Évaluation initiale rapide multi-cluster sur les inondations au Moyen Chari, Tandjilé, Mayo Kebbi Est – Tchad (tous disponibles auprès de Humanitarian Response. Pour plus d'informations : <http://assessments.humanitarianresponse.info/mira-reports>. Un quatrième rapport, disponible à la même page, n'a pas été pris en compte); Joint Rapid Assessment of the Northern Governorates of Yemen; Joint Rapid Assessment of Northern Syria – Aleppo City; et Rapid Initial Needs Assessment – Haiti (tous obtenus en utilisant le terme de recherche « assessment report » sur le site web d'ACAPS. Deux autres rapports identifiés grâce à cette recherche n'ont pas été pris en compte).

devaient évaluer la fiabilité des informateurs clés et trianguler leurs réponses les unes avec les autres, avec les sources secondaires ainsi qu'avec leurs propres observations. Lorsque des divergences inexplicables étaient constatées, les données de l'informateur n'étaient pas exploitées (Groupe de travail sur l'évaluation pour le nord de la Syrie, 2013). Une approche similaire a été adoptée en 2011 dans une évaluation au Yémen (ACAPS, 2011). Cependant, d'autres rapports examinés dans le cadre de la présente étude ne spécifiaient pas les méthodes employées pour établir la véracité des réponses des informateurs clés et des groupes cibles.

Un second et tout aussi important défi posé à l'*exactitude* des évaluations découle du fait que l'on s'en remet souvent à des évaluateurs et analystes spécialisés pour « combler les lacunes » ou synthétiser diverses sources d'information. En conséquence, les conclusions des évaluations sont parfois faussées – consciemment, mais plus souvent inconsciemment – par le biais des professionnels de l'aide sur le terrain. Là encore, la triangulation a un rôle à jouer : les orientations de la Croix-Rouge et du Croissant-Rouge préconisent de former des équipes d'évaluation diversifiées qui contrôlent les travaux des unes et des autres, tandis que les directives de l'IASC insistent sur l'importance de rendre les éventuels biais et lacunes explicites dans le rapport d'évaluation et d'expliquer comment les conclusions ont été obtenues. Dans ce contexte, il y a lieu également de rendre les évaluations publiques afin de donner la possibilité aux autorités situées en dehors du pays d'en examiner et d'en contester les conclusions. En effet, si tous les rapports pris en considération pour cette étude ont abordé la question du « biais de l'évaluateur », il est intéressant de noter que peu de rapports d'évaluation des trois dernières années sont aisément accessibles sur internet.



Il y a lieu également de rendre les évaluations publiques afin de donner la possibilité aux autorités situées en dehors du pays d'en examiner et d'en contester les conclusions.



Autre difficulté liée à l'*exactitude* des évaluations : dans les situations d'urgence, les circonstances (et les besoins qui en découlent) peuvent évoluer extrêmement vite. Il se peut qu'une évaluation peigne un tableau exact de la situation d'un village le seul jour où l'équipe se rend sur les lieux, mais qu'elle soit tout à fait obsolète au moment où le rapport est achevé. L'approche présentée dans les orientations provisoires de l'IASC vise à assurer la continuité des données d'évaluation dans le temps grâce à un processus progressif selon lequel les évaluations multisectorielles rapides sont suivies par des évaluations approfondies menées par les clusters (IASC, 2012a).

Toutefois, les clusters n'ont à ce jour réalisé que des progrès irréguliers dans l'élaboration d'orientations sur ces évaluations approfondies, lesquelles ne sont en outre pas encore menées de manière logique (Richard Garfield, communication personnelle, avril 2013). Et un cycle supplémentaire unique d'évaluations sectorielles ne résoudrait pas forcément le problème, qui semble être aussi conceptuel qu'il est pratique : la communauté humanitaire continue à considérer les évaluations comme des événements ponctuels plutôt que comme des processus continus, et de ce fait toute évaluation souffre des mêmes contraintes que le suivi des programmes (voir ci-dessous) – notamment d'un manque de financement et de soutien institutionnel.

Les régions qui connaissent régulièrement des situations d'urgence et où l'on pourrait s'attendre à ce que des stratégies d'évaluation des besoins efficaces et cohérentes aient été mises en place au cours des 20 dernières années continuent d'être mal desservies (voir Poole et Primrose, 2010 ; Slim, 2012). Et même dans les contextes (assez rares) où des évaluations séquentielles ont été menées, le recours

à des questionnaires et à des méthodes d'évaluation différents a rendu impossible toute comparaison des résultats et, en conséquence, toute compréhension de l'évolution des besoins dans le temps (Garfield et coll., 2011). La situation décrite dans l'évaluation en temps réel (ETR) de l'intervention dans la Corne d'Afrique pourrait valoir pour la plupart des opérations humanitaires : en dépit de la bonne qualité des évaluations, le système « manquait de suivi dans le temps... [les informations] n'étaient souvent plus d'actualité au moment où elles étaient traitées et lues par les HCT [équipes humanitaires dans le pays] » (Slim, 2012 : 10).

Le second élément de qualité probante examiné dans le présent rapport est la représentativité : en l'occurrence, dans quelle mesure les informations recueillies pour l'évaluation sont représentatives du groupe plus large nécessitant de l'aide. Comme nous l'affirmons plus haut, la plupart des évaluations des besoins s'appuient sur la méthode d'échantillonnage au jugé pour sélectionner les communautés, les ménages ou les individus qui participeront à l'évaluation. Les orientations issues du NATF et de l'ACAPS soulignent que, de ce fait, les conclusions des évaluations « ne sauraient représenter la totalité de la population touchée par une catastrophe », bien qu'elles puissent servir à « comprendre les préoccupations, les problèmes et les besoins les plus pressants » (IASC, 2012 : 21) et dans bien des cas constituent les meilleures données disponibles en faveur d'une intervention.

L'échantillonnage aléatoire, probabiliste – qui permet l'extrapolation statistiquement significative des résultats à l'ensemble de la population à partir d'échantillons – est plus rare dans les évaluations de situations d'urgence, bien qu'on puisse parfois le rencontrer. Pour les évaluations des besoins au Pakistan et en Haïti, on a tenté de faire appel à des techniques d'échantillonnage aléatoire (ACAPS, 2010; OCHA et NDMA Pakistan, 2012) : l'enquête pakistanaise reposait sur des données démographiques géoréférencées pour établir un échantillon, l'évaluation haïtienne s'appuyant sur l'échantillonnage spatial, c'est-à-dire l'évaluation des zones géographiques sur la base d'une grille tracée sur les zones sinistrées, sans référence à la densité ou à la distribution de la population. Ces deux cas sont révélateurs. Au Pakistan, un échantillon représentatif n'a été possible que parce que le gouvernement détenait des données démographiques assez actuelles, et parce que suffisamment d'agents formés étaient disponibles. Bien que cela ait accru la représentativité des résultats, les agents recenseurs n'ont pas pu atteindre 23 % des villages échantillonnés. En Haïti, en raison des déplacements de population – auxquels s'ajoutait une carence de données démographiques – l'échantillon, bien qu'aléatoire, n'était pas forcément représentatif (Garfield et coll., 2011). S'il serait risqué d'écarter l'échantillonnage aléatoire dans tous les contextes, « les méthodes statistiques [d'échantillonnage dans les premières phases d'évaluation] ne sont normalement pas possibles à cause des contraintes de temps et d'accès » (CICR et FICR, 2008), ainsi que du manque d'information, de fonds et de personnel qualifié.

Outre l'identification d'un groupe qui puisse représenter l'ensemble de la population – à l'aide d'échantillonnage au jugé ou probabiliste –, les évaluations des besoins doivent aussi recueillir des données séparément pour différents sous-groupes de la population, lesquels auront des niveaux et des types de besoins différents. Pour être efficaces, l'évaluation et la compréhension des besoins requièrent une connaissance des différentes situations dans lesquelles se trouvent les femmes, les enfants, les personnes âgées, les personnes handicapées et ceux qui sont socialement marginalisés (Mazurana et coll., 2011 ; Jacobsen et Furst-Nichols, 2011 ; Human Rights Watch, 2011 ; Nakagawa et Shaw, 2004).

Dans les rares cas où l'usage de méthodes statistiques est possible et souhaitable, le problème peut être résolu par la randomisation stratifiée (si des méthodes d'échantillonnage aléatoire sont employées). Lorsque des techniques d'échantillonnage au jugé sont utilisées, ACAPS préconise l'exploitation de données secondaires pour recenser divers groupes sociaux et économiques et différentes zones de moyens de subsistance. Sur cette base, les évaluations accroissent la taille de l'échantillon jusqu'à ce que la situation de tous les groupes différents ait été décrite (ACAPS, 2012b). Toutefois, bien que certaines des évaluations des besoins examinées au titre du présent état des lieux se soient manifestement efforcées d'obtenir des données complètes sur les besoins spécifiques aux groupes vulnérables – l'évaluation du Yémen, en particulier, examine les besoins spécifiques des femmes dans chaque secteur – le tableau d'ensemble demeure similaire à celui que brosse Mazurana et ses collègues : celui d'un « usage limité, ponctuel et sporadique des données ventilées selon le sexe et l'âge en phase I et II des [évaluations] » (Mazurana et coll., 2011 : 3).

Les évaluations humanitaires peuvent aussi se heurter au problème de la *pertinence* des informations qu'elles contiennent, ainsi que des indicateurs qu'elles utilisent pour identifier les besoins. Les rapports d'évaluation contiennent souvent des informations parce que celles-ci sont disponibles et non parce qu'elles tendent à prouver ou à réfuter l'existence de besoins. De même, et c'est peut-être plus inquiétant, les évaluations ont auparavant eu recours à des indicateurs qui n'avaient qu'un lien très ténu ou indirect avec la situation évaluée. Les données sur les rendements des récoltes, par exemple, ont peu de chances de nous informer sur la sécurité alimentaire dans des régions où les cultures en question ne représentent pas une importante source d'alimentation pour les populations locales. Lorsque l'on évalue, « une compréhension du contexte est primordiale pour déduire le sens... d'indicateurs, car les moyens de subsistance et les stratégies d'adaptation varient » (Catley et coll., 2008).

Le plus sérieux défi lié aux données auquel les évaluations sont confrontées se situe peut-être dans le domaine de la *clarté autour du contexte et des méthodes*. Le problème de base qui se pose est celui des suppositions qui sont faites. Les évaluations des besoins sont souvent perçues comme étant des descriptions objectives et complètes de la situation dans une région à la suite d'une catastrophe – et elles se présentent même parfois comme telles. Dans cette hypothèse, les évaluations des besoins sont perçues comme répondant à la question : « quelle est la situation ici ? » et les informations en résultant sont supposées être utiles à un vaste groupe de parties intéressées. La réalité, bien entendu, est assez différente. La plupart des évaluations des besoins répondent en fait à une variante de la question : « que devons-nous fournir, et en quelles quantités ? » et les informations sont principalement utiles aux acteurs qui les ont recueillies afin de défendre un financement et de planifier les opérations.

Ce problème de *clarté autour du contexte et des méthodes* entraîne trois difficultés, distinctes, mais liées entre elles, concernant les données probantes. La première est la tendance à supposer (au lieu de prouver) qu'un besoin existe, et que ce besoin ne peut pas être satisfait par les efforts des acteurs locaux, à savoir le gouvernement, les organisations de la société civile et la population affectée elle-même. Une telle supposition peut se solder par l'apport d'une aide qui, au mieux, est inutile et, au pire, « engendre de la passivité... sape l'initiative [et]... accentue la dépendance et l'impuissance » (Anderson et coll., 2012 : 22).

Les questionnaires d'évaluation passés ont eu tendance à être structurés autour de ces suppositions (avec des questions telles que : « combien de personnes sont privées d'abri adéquat » et « combien de ménages dans votre communauté sont aux prises avec des pénuries d'aliments aujourd'hui ? ») et ne tenaient souvent pas compte des capacités des structures locales en place à répondre à la situation. Depuis plus récemment, les directives en matière d'évaluations des besoins mettent l'accent sur la nécessité de poser des questions ouvertes qui ne supposent pas qu'un besoin existe (IASC, 2012b). Ces directives précisent en outre que le but d'une évaluation n'est pas de planifier les opérations, mais de déterminer si une opération est requise ou non (CICR et FICR 2008). Ces évolutions semblent, dans une certaine mesure, avoir influencé les pratiques d'évaluation : la majorité (mais pas la totalité) de celles qui ont été étudiées dans le cadre du présent état des lieux prenait en compte les capacités locales à répondre aux besoins. Comme il fallait peut-être s'y attendre, les évaluations menées en tandem avec les pouvoirs locaux ou organisations locales (celles du Pakistan et du Tchad) semblaient les plus solides à cet égard (OCHA, 2012 ; OCHA et NDMA Pakistan, 2012).

La deuxième difficulté concerne l'ampleur du besoin. Le fait que les évaluations forment la base des décisions de financement et ne décrivent pas de façon neutre la situation après la crise pourrait encourager les agences soit à exagérer la véritable ampleur des besoins (pour obtenir plus de fonds), soit à la minimiser (pour être « réaliste » au regard des fonds disponibles) (Darcy et Hofmann, 2003). Un exemple récent de cette attitude a été l'échec de l'évaluation des besoins en Somalie en 2011, en partie le résultat d'une « pression à rester bas, démontrant une tendance à adapter les demandes humanitaires à ce que... les bailleurs pourraient donner » (Slim, 2012 : 10).

De plus, les évaluations – qui définissent des besoins que l'État est incapable de satisfaire – constituent des actes intensément politiques, et les chiffres peuvent être systématiquement « révisés à la baisse » de ce fait (Slim, 2012 : 14) au stade de l'analyse (voir aussi OCHA ET NDMA Pakistan, 2012 ; Darcy et coll., 2013). Encore une fois, le système humanitaire semble conscient du besoin de rendre ses évaluations plus impartiales : la mise en place des ACAPS a été motivée en partie par le désir de fournir aux décideurs des évaluations indépendantes. Pour autant, il reste du chemin à parcourir avant que des évaluations indépendantes soient disponibles pour toutes les situations d'urgence : pour l'heure, « le système [continue de] manquer de données chiffrées tangibles sur les populations dans le besoin » (ALNAP, 2012 : 42).

Enfin, troisième difficulté : la compréhension des besoins. Diverses études ont montré que les perceptions locales de l'aide différaient souvent de manière marquée de celles des acteurs internationaux (se reporter notamment à Dijkzeul et Wakenge, 2010). En particulier, la compréhension et la hiérarchisation des « besoins » pour les populations touchées par des catastrophes diffèrent d'un contexte à l'autre (Anderson et coll., 2012 ; Harragin et Chol, 1998). En dépit de ces différences, la plupart des évaluations des besoins sont organisées de manière sectorielle, avec des questions à poser sur certains types de besoins prédéterminés. Conséquence : elles fournissent des données probantes liées aux besoins que les acteurs internationaux sont le mieux en mesure de satisfaire, plutôt que des preuves des besoins eux-mêmes.



la compréhension et la hiérarchisation des « besoins » pour les populations touchées par des catastrophes diffèrent d'un contexte à l'autre



De même, les définitions, imposées de l'extérieur, de « groupes vulnérables » censés avoir besoin d'aide peuvent diverger considérablement des perceptions et de l'entendement des personnes touchées par des catastrophes. Selon les conclusions du rapport d'ALNAP *L'État du système humanitaire* (2012), « dans plusieurs contextes et types d'urgence, les évaluateurs ont signalé le problème des idées préconçues sur la vulnérabilité, qui se traduisait par des interventions inappropriées » (ALNAP, 2012 : 50).

Pour résoudre cette question, certaines initiatives ont été lancées. Le DFID, par exemple, a récemment défini le devoir des humanitaires de « systématiquement impliquer les bénéficiaires dans la collecte et l'exploitation des données », notant que cela demandera peut-être de nouvelles compétences et méthodes (DFID, 2012 : 31). Entre-temps, sur le terrain, des initiatives telles que l'Échelle des besoins ressentis dans les situations humanitaires d'urgence (HESPER) visent à améliorer la compréhension des besoins. Le développement de la technologie des communications offre lui aussi aux humanitaires de bien meilleures possibilités de connaître l'opinion des populations touchées (Encadré 4).

De plus amples travaux sont cependant nécessaires. Si nous avons fait du chemin depuis le verdict accablant de la coalition d'évaluation de l'intervention à la suite du tsunami (*Tsunami Evaluation Coalition*) (Telford et Cosgrave, 2006), le rapport d'ALNAP de 2012, *État du système humanitaire*, a constaté que les enquêtes sur le terrain sur les récipiendaires de l'aide réalisées au titre du présent rapport... ont révélé que deux tiers des personnes interrogées se disaient insatisfaites ou satisfaites en partie seulement de la quantité et de la qualité du paquet global d'aide qu'elles avaient reçu », et que l'une des principales raisons était la non-compréhension du contexte local (ALNAP, 2012 ; se reporter également à Grünewald et coll., 2010).

La prise en compte dans la conception des programmes de cette compréhension des besoins au niveau local constitue une étape importante pour assurer leur efficacité (Featherstone, 2013). Toutefois, l'importance d'écouter les personnes les plus touchées et de prendre en compte leur réalité a des incidences au-delà de l'efficacité immédiate. Les résultats du CDA (*Collaborative Learning Project*) en témoignent : son projet d'écoute (« Listening Project ») visait à recueillir les conversations de plus de 6000 bénéficiaires d'aide internationale, lesquels se sont exprimés d'une seule voix.

Leur message : les évaluations conçues par les acteurs internationaux et pour eux contribuent à créer une situation où, dans les urgences, trop d'aide arrive trop vite, situation qui, selon de nombreuses personnes, satisfait les besoins institutionnels des organisations internationales plutôt que ceux ressentis sur le terrain. En même temps, la supposition selon laquelle les populations touchées sont essentiellement des récipiendaires passifs de l'aide, ce qui, comme nous l'avons vu, sous-tend l'idée derrière les évaluations des besoins, peut entraîner une dépendance à long terme, et la nécessité d'une aide continue et accrue dans le temps (Anderson et coll., 2012).

Il convient de rappeler ici que les questions que nous posons et les données probantes que nous exploitons pour y répondre ne sont pas neutres. Comme nous l'avons vu, les données probantes répondent toujours à une proposition, laquelle a été formulée par un individu à partir de certaines exigences et hypothèses. Nos choix de questions, sources et méthodes ne reflètent pas seulement le système humanitaire que nous avons, ils façonnent celui de demain.

EN BREF :

- Il n'existe pas de liste de questions d'évaluation unique pouvant être utilisée dans tous les cas, d'où une grande variété au niveau de la nature, des méthodes et de la qualité des évaluations des situations d'urgence.
- Pour résoudre ce problème, on observe de plus en plus d'initiatives visant à standardiser les méthodologies (tout en laissant de la marge pour leur individualisation) et à mener davantage d'évaluations conjointes ou coordonnées.
- Dans la plupart des évaluations, la connaissance du contexte local et le jugement individuel sont déterminants.
- Les évaluations ont tendance à être considérées comme des événements ponctuels, mais sont plus efficaces lorsqu'elles s'inscrivent dans un processus de collecte et d'analyse des données à plus long terme.
- Diverses démarches sont utilisées pour améliorer l'*exactitude* des évaluations.
- La plupart des évaluations font appel à l'échantillonnage au jugé, méthode qui est susceptible d'entraver leur *représentativité* de l'ensemble de la population.
- Un nombre réduit d'évaluations ont eu recours à l'échantillonnage probabiliste : cette méthode requiert toutefois des ressources qui souvent ne sont pas disponibles (principalement du personnel qualifié et des données de référence).
- En général, les évaluations continuent à mal représenter les besoins précis des groupes sociaux marginalisés et vulnérables.
- De nombreuses évaluations sont conçues pour répondre aux besoins spécifiques en information des organisations internationales et sont structurées autour des catégories conceptuelles créées par celles-ci. En conséquence, elles sont parfois inexactes ou peu pertinentes aux besoins réels des populations touchées.

Encadré 3 Données probantes et standards humanitaires

L'expérience de l'élaboration de standards minimums pour l'intervention humanitaire est révélatrice de la disponibilité et de la qualité des données probantes dans le secteur humanitaire.

Les équipes qui élaborent des standards humanitaires, telles que Sphère et ses standards et directives pour l'aide d'urgence à élevage (LEGS) requièrent normalement deux types de données probantes. Le premier – la « donnée de référence » – est utilisé pour mettre sur pied des indicateurs qui permettront de mesurer les résultats d'un programme par rapport au standard. Si, par exemple, un standard préconise que la population reçoive « suffisamment d'eau pour satisfaire ses besoins de base », des données probantes de référence seront requises pour recommander combien de litres d'eau par personne suffisent, dans la plupart des contextes. Ainsi les agences se servant de ce standard peuvent-elles mesurer l'effet de leur programme par rapport à un niveau minimum d'adéquation.

Le second type de données probantes concerne « ce qui est efficace ». Dans le manuel de Sphère, les données probantes de « ce qui est efficace » sont exploitées pour recommander des actions clés : « les activités suggérées et les moyens à mettre en œuvre » (Le Projet Sphère, 2011) qui, dans la plupart des cas, amèneront les agences à atteindre les standards dans leur programmation. Sur le plan des « données probantes de référence », le projet Sphère fondait certains des indicateurs – notamment ceux liés aux besoins physiologiques humains – sur des chiffres issus de recherches empiriques. À titre d'exemple, l'indicateur de 2 100 kilocalories par jour est fondé sur une série d'études regroupées par le Comité international sur la nutrition (CIN) en 1995 et adoptées par la suite par l'Organisation mondiale de la santé (PAM/HCR, 1997) pour être utilisées dans les situations d'urgence. Dans bien des cas, toutefois, ce type de donnée probante n'était pas disponible : Sphère et LEGS dépendaient tous les deux beaucoup de l'expérience directe des praticiens et des avis des experts qui examinaient et débattaient les informations disponibles et prononçaient leur jugement (Catley, 2013 ; Damerell, 2013).

Les preuves de ce qui était efficace étaient en outre difficiles à trouver. Au moment où le projet LEGS a été lancé, il n'existait aucune publication évaluée par les pairs qui traite des résultats ou des retombées des programmes concernant l'élevage dans les crises humanitaires, et les évaluations et rapports disponibles auprès des agences opérationnelles contenaient très peu d'éléments ayant valeur de preuve. En général, les évaluations étaient centrées sur les réalisations (nombre d'animaux vaccinés) plutôt que sur les résultats et les impacts (effets de la vaccination sur la morbidité due à une maladie, et impact sur les moyens de subsistance humains). Dès lors, elles ne pouvaient pas fournir de données probantes de ce qui était efficace pour traiter les problématiques humanitaires. Si une telle incapacité pourrait correspondre à des difficultés méthodologiques pour établir l'*attribution*, elle reflète aussi des failles dans la conception des projets : dans bon nombre de ceux qui étaient évalués, les résultats escomptés – vies humaines sauvées et de moyens de subsistance protégés – n'étaient pas précisés.

3.4 Données probantes émanant du suivi

Le suivi, c'est-à-dire la collecte systématique et permanente de types d'informations précis – prend des formes variées dans les opérations humanitaires. Aux fins de la présente étude, nous avons relevé quatre grandes catégories d'activité de suivi :

- le suivi situationnel, lié à l'évaluation des besoins et reflétant les propositions de type A présentées dans la section 2.2, qui décrit le contexte et indique si une intervention est requise, ou le sera ;
- le suivi de programme, lié aux propositions de type B (« ce qui est efficace »), qui porte sur le recueil d'informations relatives à la mise en œuvre et les résultats d'un programme humanitaire ;
- le suivi organisationnel, qui démontre la performance d'une organisation sur l'ensemble de ses programmes ; et
- le suivi systémique, qui examine les résultats globaux du système humanitaire international dans son ensemble.

Suivi situationnel

Dans l'idéal, les évaluations des besoins humanitaires – en particulier dans les régions où les situations d'urgence sont fréquentes et chroniques – feraient partie d'un processus plus exhaustif de suivi et de compréhension des besoins humanitaires. Plutôt que des événements ponctuels, les évaluations des besoins seraient déclenchées par un suivi situationnel continu, et signaleraient une intensification de ce suivi. Cependant, si l'on s'accorde généralement à reconnaître l'importance du suivi situationnel (voir, par exemple, ACAPS, 2012a ; CICR et FICR, 2008 ; Garfield, 2013), « les données émanant d'une évaluation [de suivi] continuent de faire cruellement défaut, ce qui se solde par des lacunes manifestes pour les interventions fondées sur des données probantes » (Darcy et coll., 2013).

Il existe en fait des systèmes de suivi situationnel : Médecins sans frontières (MSF) Belgique dirige un système de veille sanitaire en République démocratique du Congo (RDC) qui effectue son suivi par le biais d'un réseau d'antennes ; par ailleurs, en Éthiopie, un système de surveillance nutritionnelle conjoint, conçu pour détecter les foyers de malnutrition, mobilise le gouvernement et les agences internationales et nationales. Toutefois, ces systèmes de suivi tendent à privilégier des secteurs précis et sont, en tout état de cause, peu nombreux. Comme nous l'avons fait remarquer dans la section précédente, malgré une coopération humanitaire de longue haleine remontant à plus de vingt ans, le suivi systématique d'indicateurs contextuels clés n'est effectué ni au Soudan du Sud (Poole et Primrose, 2010), ni en Somalie (Slim, 2012), ni en RDC (Darcy et coll., 2013).

Cette absence de suivi situationnel semble en grande partie être un problème de culture organisationnelle et de mesures d'encouragement. Les investissements continus, à long terme dans les efforts de suivi ne sont pas très compatibles avec les délais plus courts de la programmation humanitaire, et même lorsque l'occasion s'est présentée de recueillir des données épisodiques sur quelques années, les agences humanitaires ont, dans bien des cas, constamment changé leurs méthodes et les données collectées en cours de route. En réalité, peu de tentatives ont été faites pour surveiller des situations pendant plus d'un ou deux ans.

Dans certains (rares) contextes, les programmes de « développement » ont permis la collecte d'informations et des activités de suivi susceptibles de déclencher des interventions humanitaires (Mahmood et coll., 2010). Les investissements dans des activités à plus long terme de cet ordre ont cependant plus de chances d'être réalisés lorsque l'objectif de l'aide internationale s'éloigne de l'acheminement intermittent de secours d'urgence pour privilégier le soutien régulier aux populations vulnérables. C'est le cas en Éthiopie, où « l'énorme machine de l'intervention humanitaire d'autrefois est réinventée lentement pour devenir une machine ayant vocation à apporter du secours, de l'action sociale et à réduire la pauvreté » (Darcy et coll., 2013).

Suivi de programme

Dans le même temps, le manque de suivi situationnel efficace est également lié à des faiblesses dans le second type de contrôle : le suivi de programme. Cette

question a trait au problème, en matière de données probantes, de la *clarté qui entoure le contexte et la méthode* : qu'est-ce qui est demandé, pourquoi, et pour qui ? De même que les évaluations des besoins des agences sont enclines à manquer d'objectivité dans leur mesure de l'ensemble des besoins recensés dans une situation donnée, les systèmes de suivi des programmes organisationnels ont tendance à se focaliser sur le degré de conversion réussie des moyens en réalisations, plutôt que sur la façon dont ces réalisations ont eu de l'influence sur les niveaux de besoin. Selon les calculs d'une organisation humanitaire, seuls entre 13 % et 32 % des systèmes de suivi capturaient des informations sur les effets réels des programmes sur le contexte humanitaire (Willet, 2013).



les systèmes de suivi des programmes organisationnels ont tendance à se focaliser sur le degré de conversion réussie des moyens en réalisations, plutôt que sur la façon dont ces réalisations ont eu de l'influence sur les niveaux de besoin.



Il est naturel que les bailleurs de fonds – ce sont eux qui souvent posent les questions auxquelles les systèmes de suivi sont conçus pour répondre – demandent des informations sur la façon dont leurs fonds sont dépensés. Toutefois, cette attitude donne parfois lieu à un suivi qui pour beaucoup ne vérifie pas si un programme humanitaire change vraiment les choses au niveau des besoins humanitaires (Argren, 2013). Le fait que la majorité du suivi s'effectue agence par agence exacerbe ce problème : en conséquence, il est parfois difficile de dépister les changements cumulés entraînés par la programmation humanitaire dans une région donnée.

En fait, le suivi de programme partage bien des défis en matière de données auxquels les évaluations des besoins sont confrontées : difficulté d'établir des indicateurs, de savoir si les informations recueillies sont exactes et représentatives, et de comparer différents types d'information. Le suivi de programme en diffère néanmoins sur un point important : il s'agit d'un processus continu et non d'un événement ponctuel, ce qui entraîne des défis supplémentaires, mais peut aussi améliorer la qualité des données probantes produites par les mécanismes de suivi.

Nous avons constaté que l'accès physique à des milieux dangereux rendait difficile le recueil d'un grand nombre d'informations humanitaires. Ce défi est particulièrement prononcé pour le suivi de programme, qui requiert un accès fréquent et régulier aux opérations. Or, un accès limité rend extrêmement difficile d'être sûr de l'*exactitude* ou de la *représentativité* de toute information et d'ensembles de données accumulées au fil du temps.

L'amélioration des télécommunications peut aider les agences à accéder à des informations en provenance de zones précaires (voir encadré 4). En outre, les agences font appel à des solutions « humaines » innovantes pour résoudre ce problème. Dans l'est du Myanmar, le Comité international de secours (IRC) s'est joint à d'autres agences pour créer un système par lequel, chaque fois qu'une agence a accès à une région, elle suit les projets de toutes les agences de cette région. Cette approche atténue le risque de biais de la part du contrôleur (et améliore donc l'*exactitude*), tout en augmentant la taille de l'échantillon, ce qui permet d'effectuer une plus grande triangulation et davantage de contrôles pour vérifier que les conditions surveillées sont représentatives. En Somalie, l'UNICEF a abordé un problème similaire de manière assez différente : elle a utilisé la validation par un tiers pour garantir la qualité de l'information en provenance des systèmes de suivi. Lorsque des contrôles qualité de base donnent à penser qu'il pourrait y avoir

des problèmes au niveau du suivi de l'information, l'UNICEF envoie des contrôleurs tiers qui se chargent d'une enquête de suivi. Ces contrôleurs remplissent leur mission « en aveugle », ce qui les empêche de fabriquer l'information. Là encore, cette démarche sert à améliorer radicalement l'*exactitude* des informations (Fetouh et coll., 2013).

Si le suivi humanitaire a tendance à inclure la population affectée moins que certaines autres phases du cycle de programme (Alexander et coll., 2013), il peut néanmoins l'inclure de manière plus efficace. Le fait que les activités de suivi s'inscrivent dans la durée et ont un caractère permanent rend possible une inclusion plus significative, et moins symbolique : on dispose de temps pour nouer des relations et instaurer un dialogue plus authentique que celui qui a tendance à s'établir dans les évaluations ou appréciations des besoins. La croissance récente des mécanismes de retour d'information et de doléances, qui constituent souvent un élément du suivi de programme, a conduit beaucoup d'organismes

Encadré 4 Technologie et données probantes

Grâce aux avancées réalisées dans les technologies de l'information, il est possible d'améliorer considérablement la capacité des humanitaires à recueillir et analyser des données et à assurer que celles-ci sont exploitées.

Les technologies de l'information permettent de collecter les données rapidement après la survenue d'une crise, et de le faire régulièrement au fur et à mesure que celle-ci évolue. Les TIC peuvent également rendre les données probantes plus accessibles et plus faciles à interpréter. Pourtant, faute de systèmes pouvant adéquatement traiter ces données, ces technologies peuvent être difficiles à gérer. Les humanitaires doivent renforcer leur capacité à analyser, comprendre et exploiter les données réunies. Lorsque les ressources sont maigres ou la capacité limitée, les options faiblement techniques peuvent s'avérer plus efficaces.

Tandis que les technologies continuent de développer leur énorme potentiel au service d'une action humanitaire informée par des données probantes, plusieurs autres facteurs devront être pris en considération. Si des quantités importantes de données peuvent être réunies rapidement, elles ne constituent pas toutes des données probantes utiles. Dans un même temps, la quantité accrue de données recueillies peut susciter des attentes que les humanitaires risquent de ne pas pouvoir satisfaire. De plus, une dépendance accrue de la technologie peut rendre les organismes vulnérables aux défaillances technologiques communes dans les contextes post-catastrophe.

Une des préoccupations majeures concerne la qualité de preuve des informations obtenues par crowd-sourcing, c'est-à-dire obtenues auprès de nombres élevés de personnes à l'aide de la téléphonie mobile ou de services en ligne. Les avancées technologiques rendent possible la collecte de données auprès d'acteurs nouveaux et variés. D'un côté, cela facilite la triangulation des informations et permet de broser un tableau complet ; de l'autre, l'*exactitude* des données et la *crédibilité* de certaines sources peuvent être remises en question. Par ailleurs se posent des questions de représentativité. Si les données sont rassemblées par le biais de SMS, par exemple, la source des informations est limitée à ceux qui ont accès à des téléphones mobiles.

Le projet de cartographie d'Haïti d'Ushahidi a fourni une prise de conscience situationnelle précise qui, selon une évaluation, a été particulièrement utile pour la réponse initiale. Pourtant, peu d'humanitaires ont tenu compte de cette information. Ils se méfiaient de ce crowd-sourcing et étaient inquiets de la qualité de preuve et de la représentativité. Ces données ne convenaient pas aux exigences d'information rigoureuses des organismes intervenants. Beaucoup d'entre eux craignent que les données en provenance de plateformes de crowd-sourcing soient biaisées ou manipulées. D'autres font observer que toutes les données contiennent un biais inhérent, et que la prise de décision humanitaire s'appuie fréquemment sur des données empiriques. Parfois, des données partiales valent mieux que pas de données du tout.

En améliorant le flux d'informations et la communication dans les deux sens, la technologie a le potentiel d'améliorer la redevabilité envers les populations affectées ; en fournissant des chaînes d'information claires, c'est la redevabilité envers les financeurs qu'elle encourage. Elle peut rendre la collecte de données probantes plus participative et plus facile à mener à grande échelle. Toutefois, elle soulève aussi des inquiétudes au sujet de la protection des données. Le géomarquage, c'est-à-dire l'identification géographique d'informations ou d'images, peut s'avérer utile dans l'analyse, mais il comporte des risques non négligeables de sécurité et de confidentialité. De même, si l'archivage numérique permet de gagner du temps et de réduire la duplication d'erreurs, beaucoup d'organismes appréhendent de stocker de vastes quantités de données dans des serveurs ou dans des systèmes basés sur le nuage. Les humanitaires doivent élaborer des directives pour assurer que les informations sont utilisées de manière éthique et sûre.

Le développement de plateformes communes constitue un autre domaine où l'action à l'échelle du système pourrait être utile. Les données probantes sont les plus utiles lorsqu'elles sont aisément accessibles, et si le large éventail d'applications de données customisées rend les données plus facilement compréhensibles à leurs usagers, il est d'une utilité limitée pour les personnes qui utilisent un autre système ou une autre plateforme.

Les travailleurs humanitaires doivent reconnaître les perspectives que les nouvelles technologies leur offrent pour améliorer la qualité et l'utilité des données probantes et contribuer activement à assurer que, à mesure que ces technologies sont mises en œuvre, elles abordent les questions de qualité, confidentialité, accessibilité et interopérabilité des données.

Sources : Avila, R. et coll. (2010) ; Coyle, D. & Meier, P. (2009) ; OCHA (2013b) ; Smith, G. et coll. (2011) ; World Vision (2013) ; FICR (2013)

à expérimenter une multitude de voies de retour, en vue de mettre en place des systèmes robustes, propres au contexte et permettant aux groupes marginalisés de fournir des informations sur leur expérience de l'assistance (Bonino et Jean, à paraître ; Fenton et coll., 2011). Comme dans l'évaluation des besoins, on a déjà la preuve que ces mécanismes améliorent le caractère probant des informations de suivi (Featherstone, 2013 ; Laybourne et Obrecht, 2013).

Suivi organisationnel

Outre le suivi de programmes individuels, un nombre croissant d'organismes humanitaires de l'ensemble du secteur essaient de contrôler leur performance globale en tant qu'acteurs humanitaires en consolidant les informations recueillies à l'échelon des pays ou des programmes, et ce afin d'obtenir un tableau d'ensemble. La nature de ce qui est contrôlé diffère d'un organisme à l'autre : ainsi, Mercy Corps examine la performance par rapport à la mission de l'organisation (Willet, 2013), Action contre la faim (ACF) par rapport aux critères d'évaluation de l'OCDE-CAD (Guerrero et coll., 2013) ; l'association Catholic Relief Services (CRS), elle, privilégie surtout les mesures des réalisations (Carr, 2013). Ces différences correspondent à la nature des données recueillies et au degré d'inclusion de données supplémentaires (outre celles fournies par le suivi de programme et les systèmes d'évaluation). Si Mercy Corps a très clairement expliqué que son système « paramètres de mission » n'exploiterait que des données en provenance de systèmes de suivi existants, Oxfam GB augmente ses rapports de programme d'entretiens avec le personnel des programmes (Walden, 2013).



Il est extrêmement difficile d'établir des catégories communes de mesures sur l'ensemble des programmes.



Ces initiatives rencontrent certaines difficultés communes en matière de données probantes. En raison des contextes très différents dans lesquels opère un organisme donné, et de la grande diversité des programmes qu'il peut mettre en œuvre, il est extrêmement difficile d'établir des catégories communes de mesures sur l'ensemble des programmes (Carr, 2013). D'ailleurs, même le terme « bénéficiaire » peut revêtir des sens différents selon les contextes. Lorsque les informations proviennent d'un sous-ensemble de programmes, des problèmes de représentativité peuvent également se poser. Dans quelle mesure ces programmes représentent-ils l'action de l'organisme concerné dans son ensemble ?

Comme pour les autres domaines étudiés, le défi le plus fondamental est peut-être celui de la *clarté autour du contexte et des méthodes* : de quelle manière les informations sont-elles censées être utilisées, et par qui ? Un spécialiste avance : « La question 'Où en sommes-nous ?' ne constitue pas une question de recherche valable » (Willet, 2013). Les agences produisent de vastes quantités d'informations dont l'analyse requiert des ressources qui pourraient être affectées ailleurs. Le suivi de la performance organisationnelle peut avoir plusieurs usages : développement stratégique, plaidoyer, rapports aux donateurs, et – lorsque les données sont ventilées par pays – gestion des performances. Compte tenu des implications de ces approches pour les ressources, il y a lieu de mettre sur pied des systèmes de suivi organisationnel pour répondre à des questions claires et explicites.

Correctement effectué, ce contrôle des performances à l'échelle de l'organisme peut promouvoir une amélioration dans les organisations individuelles. Pour autant, l'accroissement de la taille, de la portée et de la complexité des interventions humanitaires au cours des dix dernières années a entraîné une interaction et une interdépendance accrue entre les organismes humanitaires, qui tentent de répondre à des situations nécessitant des interventions à grande échelle, à phases multiples et multisectorielles. Dans un tel contexte, le suivi et l'amélioration des performances des organismes individuels constituent une importante première étape vers une meilleure intervention humanitaire. Dans l'idéal, cependant, il faudrait pouvoir contrôler la performance de l'ensemble du système d'assistance humanitaire.

Suivi systématique

Il existe divers cadres et systèmes pour contrôler les performances humanitaires. Ils tendent à se concentrer sur les performances de certains groupes d'acteurs au sein du système humanitaire plus large. C'est le cas de l'initiative Good Humanitarian Donorship, qui s'attache à évaluer les performances des bailleurs de fonds par rapport à des principes précis (voir DARA, 2011). D'autres, tel l'Humanitarian Accountability Partnership, se concentrent sur les performances liées à certains éléments spécifiques de l'intervention humanitaire, en l'occurrence la redevabilité envers les personnes affectées (voir HAP, 2010) ; d'autres encore s'intéressent aux performances liées à des réponses précises, par exemple la réponse internationale au tsunami dans l'océan Indien (Cosgrave, 2007).

Jusqu'à récemment, « aucune tentative n'a néanmoins été faite d'évaluer la performance du secteur humanitaire dans son ensemble, soit de manière ponctuelle, soit régulièrement » (Ramalingam et coll., 2009 : 73). Aucun cadre convenu commun n'existe pour surveiller les performances humanitaires internationales, et bien que les mécanismes spécifiques de suivi et d'élaboration de rapports abondent, comme nous l'avons vu, ils sont fragmentés et recueillent toutes sortes d'informations pour différentes raisons.

En 2009, dans le cadre du Projet sur les performances humanitaires du Réseau, ALNAP avançait qu'« un cadre équilibré, complet et cohérent au service de la performance humanitaire pourrait être discuté et adopté dans toutes les agences » (Ramalingam et coll., 2009 : 7) et proposait un « tableau de bord humanitaire équilibré » en guise de base de surveillance à l'échelle du système. À ce jour, l'idée d'un cadre de rendement commun n'a pas fait beaucoup de chemin, bien que le sujet ait connu un regain d'intérêt lors de discussions récentes, liées à la proposition d'un sommet humanitaire mondial en 2016, sur le thème de l'efficacité humanitaire (OCHA, 2013). Entre-temps, ALNAP a instigué un processus de suivi des performances du système humanitaire par rapport aux critères de l'OCDE-CAD¹⁰ avec son rapport bisannuel *L'État du système humanitaire* (ALNAP, 2010 et 2012). Ce processus vise à décrire l'ampleur, la morphologie et les résultats du système humanitaire officiel, à identifier des tendances dans ses performances, et – peut-être plus important encore – à préciser les grandes problématiques et les lacunes que des efforts d'amélioration du secteur devraient combler.

Si ce rapport a été extrêmement bien accueilli (John Mitchell, communication personnelle, juin 2013 ; Registre du suivi et de l'évaluation d'ALNAP), il met aussi en lumière une série de difficultés liées à la collecte et à l'analyse des données (notamment celles portant sur les propositions de type B : « ce qui est efficace ») provenant d'une multitude d'agences et d'organismes opérant dans une multitude de contextes.

Le premier des ces défis renvoie aux difficultés relevées dans les sections sur l'alerte précoce et l'évaluation, à savoir la *pertinence* des indicateurs de mesure du rapport. Les critères de l'OCDE-CAD utilisés dans le rapport, tels que ceux de « pertinence » et de « cohérence », sont des catégories conceptuelles comportant un élément subjectif qui, dans la plupart des cas, ne peut être mesuré qu'à l'aide de divers

10 Pour plus d'informations sur les critères d'évaluation de l'OCDE-CAD appliqués à l'action humanitaire, voir Beck (2006).

types de données. Les choses se compliquent davantage encore : il n'existe en effet aucun accord général sur quels jeux d'indicateurs pourraient ou devraient être retenus pour mesurer chaque critère (et, comme on l'a mentionné, il est peut-être impossible d'établir un jeu unique d'indicateurs, compte tenu de la variété des contextes où opèrent les humanitaires).

Dès lors, la sélection des indicateurs pour le rapport dépend beaucoup de l'opinion d'experts. ALNAP adopte une démarche en trois étapes en vue de diminuer tout biais éventuel dans l'analyse. La première mouture du rapport est préparée par des chercheurs spécialisés qui maîtrisent bien le système. Cette version est analysée en détail par le secrétariat d'ALNAP, puis examinée séparément par un Groupe consultatif indépendant composé d'individus issus du réseau ALNAP ou situés en dehors de celui-ci et qui apportent des perspectives diverses. La diversité des contextes représentés par le Groupe consultatif permet aussi d'assurer que les questions difficiles à mesurer ne sont pas exclues du rapport final (Knox Clarke, 2013a).

Le deuxième défi – à rapprocher de l'*exactitude* des informations et de l'analyse – ne nous est pas inconnu : c'est celui de la carence même des données. Pour plusieurs domaines sous surveillance, nous ne disposons que de maigres données convaincantes. Par exemple, s'agissant de la couverture, « le système manque de données chiffrées concrètes sur les effectifs totaux ayant besoin d'aide humanitaire à un moment donné » (ALNAP, 2012 : 42). Lorsque des données existent, elles sont produites en général par des acteurs issus de l'intérieur même du système, et sont donc susceptibles d'être biaisées. Dans une certaine mesure, ce biais peut certes être éliminé par triangulation, mais il est important d'effectuer cette triangulation avec des informations émanant d'acteurs situés en dehors du « système formel » afin de contester éventuellement les biais systémiques ou préjugés culturels pouvant exister au sein d'un système global. À ce titre, l'inclusion d'enquêtes auprès des bénéficiaires et d'entretiens avec le personnel des agences nationales de gestion des catastrophes constitue un élément décisif de la méthodologie de recherche.

Comme on l'a mentionné, toute mesure de la performance globale du secteur humanitaire – que ce soit au niveau de l'agence individuelle ou de l'ensemble du système – sera confrontée à des difficultés de *représentativité* : dans quelle mesure l'échantillon d'information reflète-t-il correctement toute la situation ? Dans les éditions précédentes de *L'État du système humanitaire*, les analyses des évaluations et des entretiens avec les informateurs clés reposaient sur des échantillons fondés sur l'objet visé, tandis que les visites de terrain (de « vérification sur place ») et les enquêtes auprès des bénéficiaires se fondaient sur des échantillons de commodité. Dans la prochaine phase du projet, ces visites et ces enquêtes seront sélectionnées à l'aide d'une démarche fondée sur l'objet visé, dans un effort de représenter plus étroitement l'ensemble de l'action humanitaire à travers les informations découlant de ces sources (John Mitchell, communication personnelle, juillet 2013).

EN BREF :

- Le suivi est composé d'activités séparées diverses mais liées entre elles : suivi situationnel, suivi de programme, suivi organisationnel et suivi systémique.

- Le système humanitaire est faible en matière de suivi situationnel : peu de tentatives réussies ont été faites pour surveiller des situations sur des périodes de plus d'un ou deux ans.
- Le suivi des performances a tendance à être trop axé sur les réalisations et pas assez sur les résultats et les impacts ; il souffre par ailleurs de problèmes d'*exactitude*, notamment dans les situations où l'accès est difficile.
- Les agences ont mis au point une palette de moyens novateurs pour s'attaquer à ces problèmes. Plusieurs organismes du réseau d'ALNAP tentent de contrôler leur rendement humanitaire global. Le suivi organisationnel – en particulier lorsqu'il exploite les informations existantes – est difficile, mais la clarté sur l'objet à contrôler et la façon dont l'information sera utilisée est essentielle à son succès.
- ALNAP surveille la performance de l'ensemble du système humanitaire dans son rapport bisannuel *L'État du système humanitaire*, mais le défi est difficile à relever, d'une part parce beaucoup des données nécessaires n'ont pas été recueillies, d'autre part à cause de difficultés à formuler des généralisations à partir d'un éventail très varié de situations.

3.5 Données probantes émanant des évaluations et des essais contrôlés



Les évaluations ont plus de chances de fournir des données probantes solides lorsqu'elles font appel à la démarche dite des méthodes mixtes.



Les évaluations de l'action humanitaire sont confrontées à des défis similaires aux appréciations des besoins et au suivi. Le lecteur connaît d'ores et déjà ces défis : les évaluations ont lieu dans des milieux pauvres en données, politisés et complexes, physiquement d'accès limité, où les populations sont mobiles et où évoluent divers acteurs, lesquels souhaitent tous donner leur point de vue sur les événements. De tels contextes ne facilitent pas la constitution d'un corpus de données qui soit exact, représentatif de l'expérience des populations touchées et non faussé par des interprétations subjectives.

En général, les orientations en matière d'évaluation de l'action humanitaire tendent à indiquer que les évaluations ont plus de chances de fournir des données probantes solides lorsqu'elles font appel à la démarche dite des méthodes mixtes. À titre d'exemple, les directives de la Fédération internationale des sociétés de la Croix-Rouge et du Croissant-Rouge (IFRC) en matière de suivi et d'évaluation indiquent que les données qualitatives ne permettent qu'un degré limité de *généralisation* et peuvent être perçues comme peu crédibles, alors que les méthodes quantitatives peuvent être coûteuses et « exclure les commentaires et les voix humaines expliquant pourquoi quelque chose s'est produit ». En conséquence, « une approche mixte est souvent préconisée, car elle met à profit les avantages des deux méthodes » (IFRC 2011 : 35). De même, l'unité évaluation de MSF affirme que « normalement, un mélange de méthodes qualitatives et quantitatives produit les meilleurs résultats » (MSF 2012 : 7), tandis que selon les directives du Programme alimentaire mondial (PAM), « dans la mesure où les données qualitatives et quantitatives sont complémentaires, les deux devraient être utilisées » (PAM, s.d. : 23).

Dans l'esprit de ces directives, le PAM et le Haut-Commissariat des Nations unies pour les réfugiés (HCR) ont récemment achevé une série d'évaluations d'impact à méthode mixte sur la contribution de l'assistance alimentaire à des

solutions durables dans les situations de réfugiés prolongées (Steen et coll., 2013). Ces évaluations utilisaient des données quantitatives émanant d'enquêtes auprès des ménages¹¹, fondées sur des échantillons randomisés de la population de réfugiés, et des données qualitatives issues d'entretiens, de groupes cibles et de l'observation. Le rapport de synthèse constate que « la principale garantie de validité des données recueillies est la large palette des sources et des méthodes employées pour rassembler les données probantes et leur triangulation » (Steen et coll., 2013 : 3). L'échantillonnage probabiliste – qui améliore la représentativité des données – a été également utilisé dans des évaluations en milieux pastoraux lorsque des listes de ménages étaient disponibles (Abebe et coll., 2008).

En général, toutefois, l'évaluation humanitaire « fait principalement appel aux méthodes qualitatives » (Buchanan-Smith et Cosgrave, 2012). Selon un bilan informel réalisé dans le cadre de l'étude sur les évaluations de 2010-2012 disponible dans le Centre de ressources d'ALNAP, l'usage des méthodes mixtes est inhabituel, et la plupart des évaluations ont misé sur des démarches globalement qualitatives pour produire des preuves, et notamment sur des entretiens – souvent avec des informateurs clés – et l'observation personnelle.

La plupart des évaluations ont recours aux techniques d'échantillonnage au jugé pour identifier les personnes interrogées. Pour établir l'*exactitude*, elles se fondent sur la triangulation des sources (et, dans une certaine mesure, sur la triangulation des observations des différents évaluateurs). L'orientation vers des approches qualitatives et discursives est particulièrement marquée dans les évaluations dont la première vocation est l'apprentissage (plutôt que la redevabilité). En effet, ce type d'évaluation met l'accent sur l'importance de l'expérience subjective et sur la participation des parties prenantes clés au processus d'évaluation comme conditions préalables à tout apprentissage et changement.

Si l'on peut parfois être enclin à voir les approches quantitatives de la production de données probantes comme « tangibles » et les approches qualitatives comme « intangibles » d'une certaine manière, l'usage de ces dernières ne doit pas être prétexte à un manque de rigueur. Les six critères présentés à la section 2.3 pour définir une donnée probante solide sont aussi importants pour les informations glanées à travers les entretiens que pour les données issues des enquêtes randomisées. Malheureusement, maintes évaluations humanitaires n'emploient pas de méthodes qualitatives reconnues par les milieux de la recherche ; elles ne parviennent pas à satisfaire les normes de qualité de base relatives à l'*exactitude*, à la *représentativité* et à la *pertinence*. L'un des organismes membres d'ALNAP a récemment analysé ses propres évaluations à la lumière des *Evidence Principles* de l'association Bond, liste de principes sur la qualité des données destinés aux ONG (Bond, s.d.). Leurs conclusions indiquent qu'il y a d'importantes améliorations à apporter :

« Selon l'instrument de Bond, la qualité des données dans les... évaluations était jugée basse dans presque toutes les catégories recensées, c'est-à-dire que

11 Mesure de la sécurité alimentaire au moyen des indicateurs communs du score de consommation alimentaire (SCA), du score de diversité alimentaire des ménages (HDDS) et de l'Indice de stratégie d'adaptation (ISA).

la qualité des projets évalués était bonne, mais qu'il était difficile de tirer cette conclusion à partir des rapports en raison de la mauvaise qualité des données. De nombreuses conclusions ne semblaient reposer sur aucune donnée probante, et beaucoup d'anecdotes étaient fournies en guise d'illustrations de conclusions plus générales. Les rapports contenaient des expressions d'opinions, mais des exemples ou éléments probants étaient rarement donnés pour les justifier. On constatait souvent un manque d'informations quantitatives de base (par exemple, le nombre d'hommes, de femmes, de personnes âgées ou de jeunes rencontrés) et une absence de méthode analytique ou de méthodologie convaincante. » (Parker et Sanderson, 2013).



Autre sujet de préoccupation dans beaucoup d'évaluations : le manque de points de vue des bénéficiaires.



Certes, ces constatations ne sont peut-être pas représentatives de toutes les organisations humanitaires, mais – là encore, si l'on se fonde sur le bilan informel du Centre de ressources d'ALNAP – elles ne semblent pas particulièrement inhabituelles.

Mesurant l'importance d'assurer la qualité probante des évaluations, plusieurs organismes ont ces dernières années mis en place des systèmes pour examiner la qualité de leurs évaluations : le PAM est doté d'un tel système¹², et OCHA projette d'en créer un (Jock Paul, communication personnelle, avril 2013). Parallèlement, ALNAP expérimente actuellement un Guide pour l'évaluation de l'action humanitaire qui définit les grandes lignes de la mise en œuvre de méthodes qualitatives plus robustes (Buchanan-Smith et Cosgrave, 2013).

Autre sujet de préoccupation dans beaucoup d'évaluations : le manque de points de vue des bénéficiaires, malgré le recours massif aux entretiens comme source d'information évaluative. Lorsque Beck et Buchanan-Smith ont procédé à des méta-évaluations pour ALNAP, ils ont découvert que près des trois quarts des évaluations examinées entre 2001 et 2004 n'avaient pas consulté les bénéficiaires, ou n'avaient accordé qu'une part minimale à la consultation (Beck et Buchanan-Smith, 2008).

En dépit de quelques exceptions notables, les évaluations tendent encore à sous-estimer l'expérience des populations affectées en tant que source de données : ainsi, l'édition de 2012 de *L'État du système humanitaire* (ALNAP, 2012) conclut-elle que la consultation des récipiendaires est l'un des points les plus faibles de la performance humanitaire, tandis que le rapport sur la redevabilité humanitaire de l'Humanitarian Accountability Partnership (HAP) fait observer que la coopération est « extrêmement rare dans l'évaluation, même si la tendance actuelle est d'encourager l'implication des bénéficiaires à ce stade » (Alexander et coll., 2013 : 29). Les méthodes participatives d'impact décrites ci-dessous nous fournissent un exemple intéressant, quoiqu'atypique, de la manière dont l'évaluation peut être bâtie autour de l'expérience des personnes touchées par des catastrophes.

12 Le Système d'assurance qualité des évaluations (EQAS selon le sigle anglais), qui énonce un certain nombre de processus intégrant des mesures d'assurance qualité et des modèles de produits d'évaluation. Il comprend en outre une assurance qualité des rapports d'évaluation (rapports initiaux, complets et de synthèse) à partir de listes de contrôle standardisées... Ce processus d'assurance qualité ne nuit pas à l'indépendance des opinions de l'équipe d'évaluation, mais assure que le rapport fournit les données requises de manière claire et convaincante et tire ses conclusions sur cette base. » (FAO et PAM, 2013).

Causalité et transférabilité – deux enjeux particuliers pour les évaluations

Outre les contrôles de la nature probante des données décrits dans la section 2, les évaluateurs doivent relever le défi de l'*attribution*. Une évaluation ne doit pas se contenter de décrire exactement une situation ; elle doit aussi montrer le lien entre une intervention précise, ou série d'interventions, et la situation décrite (propositions de type B). Dès lors, les évaluations doivent être rigoureuses dans leur approche de l'*attribution*.

Bon nombre d'évaluations de l'action humanitaire abordent ce défi de *causalité* en s'appuyant sur un cadre logique. En vertu de cette démarche, un projet est conçu selon un enchaînement causal, qui forme une sorte d'hypothèse. Si un certain nombre d'objectifs à atteindre sont réalisés (par exemple, un certain nombre de puits produisant une quantité et une qualité d'eau spécifiées) et si certaines suppositions sont vérifiées (les gens se procurent leur eau potable à cette source et non ailleurs), alors la supposition « logique » est que l'on observera certains résultats positifs, par exemple une baisse des maladies d'origine hydrique. Dans ce cas, si une évaluation peut démontrer que les objectifs à atteindre ont été réalisés et que les résultats en ont découlé, et si les personnes interrogées créent un lien narratif entre les objectifs et les résultats qui écarte toute autre explication, alors l'objectif à atteindre est généralement considéré comme ayant causé le résultat.¹³

Cette démarche, bien qu'imparfaite, a tendance à prévaloir chaque fois que les évaluations sont axées sur le niveau de résultats. Cependant, elle est moins utile lorsque l'évaluation prend en compte l'impact de l'intervention humanitaire, parce que l'enchaînement causal entre objectif à atteindre et impact¹⁴ a tendance à être plus complexe et plus ambigu et n'a peut-être pas été formulé auparavant.

La mesure de l'impact de l'action humanitaire est difficile pour plusieurs raisons : manque de capacité ; forte rotation des effectifs ; aversion à divulguer un échec ; enfin, difficultés techniques à établir des groupes de référence et de contrôle et à dissocier l'impact d'une seule intervention de l'impact plus large d'une opération. Autant de causes qui vont à l'encontre d'une évaluation d'impact robuste. D'où la relative rareté des évaluations d'impact dans le secteur humanitaire (Proudlock et Ramalingam, 2009). Lorsqu'elles sont réalisées, le sens du mot « impact » diffère souvent d'une agence à une autre (ALNAP, 2012 ; ACF, Mimeo). Il faut noter toutefois un intérêt accru récent et considérable pour l'évaluation de l'impact humanitaire, motivé – en partie du moins – par une volonté de déterminer quelles approches offrent le meilleur rapport coût-résultats dans un contexte financier austère.

Au-delà du secteur humanitaire, on assiste à un débat animé sur les approches les plus intéressantes pour établir des données probantes solides d'*attribution*. Les défenseurs de démarches expérimentales – les essais cliniques randomisés (ECR), en général – affirment qu'elles représentent « l'étalon d'or » pour établir la causalité ; exprimé de manière plus modeste, « la pire forme de conception à l'exception de

13 Il est intéressant de noter que cette approche s'apparente à la définition rationaliste plutôt qu'empirique de la vérité, puisqu'elle s'appuie sur une séquence logique plutôt que sur l'observation directe d'un lien causal.

14 Souvent interprété comme le lien final dans un enchaînement causal : « Effets à long terme positifs et négatifs, primaires et secondaires produits par une intervention de développement, directement ou indirectement, de manière intentionnelle ou non » (OCDE-CAD, 2002).



Le nombre d'essais contrôlés, de bilans systématiques et d'autres démarches qui privilégient des méthodologies expérimentales devrait augmenter, mais dans de nombreuses situations, les démarches expérimentales ne sont ni souhaitables, ni possibles.



toutes les autres méthodes qui ont été essayées » (Bickman et Reich, 2009). Les critiques pointent le coût des ECR – normalement entre 200 000 et 900 000 dollars (Banque mondiale, s.d., cité dans Bradt, 2009) – faisant observer que des démarches moins coûteuses, non expérimentales sont employées régulièrement dans toute une gamme de disciplines scientifiques afin d'établir la causalité sans le moindre doute (Scriven, 2009). Selon eux, les résultats des ECR ne sont vraiment efficaces que lorsque les interventions sont « stables et relativement simples et... produisent des effets relativement rapides et conséquents par rapport à d'autres influences potentielles » (Piccioto, 2012). Les critiques maintiennent par ailleurs que les ECR sont d'une utilité limitée dans l'élaboration de politiques parce qu'ils ne sont pas généralisables à d'autres contextes (Schwandt, 2009) et parce qu'ils expliquent rarement comment une intervention a eu des incidences précises (Piccioto, 2012).

Peut-être faut-il conclure que les ECR sont certainement utiles lorsqu'ils répondent au type de question clinique précise dit PICO pour lequel ils ont été conçus (Bradt 2009).¹⁵ Au-delà de la sphère clinique, ils ont tendance à être les plus efficaces à établir l'*attribution* « lorsque l'enchaînement causal entre l'agent et le résultat est assez court et simple et lorsque les résultats peuvent être extrapolés sans risque à d'autres milieux » (Victoria et coll., 2004, cité dans Dijkzeul et coll., 2013).

Quelques initiatives d'utilisation des ECR dans le champ humanitaire ont été tentées pour fournir des preuves d'incidences. Le DFID a financé un ECR au Malawi afin de tester les différentes compositions d'aliments prêts à être consommés dans le traitement de la malnutrition aiguë sévère (Kerak et coll., 2009 dans Buchanan-Smith et Cosgrave, 2012), tandis qu'ACF a mené des ECR sur les incidences qu'avait l'ajout d'aliments prêts à l'usage aux programmes d'alimentation complémentaire pour la malnutrition infantile (Huybregts et coll., 2012). Au Liberia, l'IRC a réalisé une évaluation d'impact sur un programme de reconstruction axé sur la communauté à l'aide d'un concept expérimental (Fearon et coll., 2008) : un exemple intéressant – et à ce jour plutôt inhabituel – de l'usage d'approches expérimentales dans l'action humanitaire hors des secteurs de la santé et de la nutrition.

Le nombre d'essais contrôlés, de bilans systématiques et d'autres démarches qui privilégient des méthodologies expérimentales devrait augmenter, avec le soutien d'organismes tels que 3IE et EvidenceAid, qui tentent d'accroître la rigueur et la sophistication de la production de données probantes dans le secteur humanitaire. Au fur et à mesure que ces travaux progressent, nous aurons beaucoup à apprendre des premières tentatives. L'expérience d'ACF au Tchad confirme que les ECR en contextes humanitaires exigent une longue période de conception et de mise en œuvre et sont à la merci de l'évolution de la situation des groupes qui y prennent part. Leur succès dépend du haut niveau de compétence des chercheurs (en l'occurrence, un partenariat avec l'université de Gand et l'Institut de médecine tropicale d'Anvers) et d'une communication fréquente et efficace avec les participants, l'ensemble de la population et le personnel de l'organisation sur le terrain (Puett et Salpeteur, 2013). En raison des coûts élevés qu'ils impliquent,

15 PICO signifie Patient (ou population) recevant une intervention ; Intervention envisagée ; Comparaison (l'intervention alternative à l'étude) ; Résultats cliniques visés. L'étude de ces quatre facteurs entraîne la création d'une question « vérifiable ». Les questions de type PICO sont souvent particulièrement utiles pour répondre à certaines questions, par exemple déterminer la thérapie la plus efficace.

il convient de sélectionner des sujets hautement stratégiques, c'est-à-dire pour lesquels le coût des dépenses est compensé par le coût de *ne pas* effectuer ces recherches (Chloe Puett, communication personnelle, avril 2013).

Pour autant, dans de nombreuses situations, les démarches expérimentales ne sont ni souhaitables, ni possibles. Stern, dans un rapport récent pour le DFID (Stern, 2012) et Rogers, dans un rapport pour Interaction (Rogers, 2012) proposent des conceptions et des méthodes alternatives, notamment les approches dites quasi expérimentales, les approches individualisées et les démarches fondées sur la théorie. Les deux auteurs soulignent l'intérêt de faire intervenir diverses approches pour étudier la causalité. Ils notent en outre l'importance, dans le contexte humanitaire, d'examiner dans quelle mesure les interventions contribuent aux changements, au lieu d'essayer d'attribuer un changement à une intervention unique.

Afin d'aborder la question de l'*attribution* (encadré 5), les humanitaires expérimentent de plus en plus avec une gamme de méthodes. Les évaluations d'impact du PAM et du HCR citées plus haut faisaient appel à des groupes témoins « opportunistes », groupes officiels et non officiels de réfugiés ou camps qui recevaient des quantités d'assistance différentes. Faute de cadre logique formel pour le programme, les évaluateurs ont également reconstruit la logique des interventions afin de voir si l'intrant (l'aide alimentaire) causait bel et bien l'incidence de la manière envisagée à l'origine.¹⁶ Le projet ECB a mis au point et testé une méthodologie qui examine la contribution des interventions humanitaires aux changements ; pour ce faire, il utilisait des paramètres descriptifs alliés à des données issues d'entretiens (Few et coll., 2013).

Comme nous l'avons vu, l'évaluation d'impact est l'un des rares domaines où des efforts concertés ont été tentés pour construire l'évaluation autour des perceptions des populations affectées. Le Feinstein Institute de l'université Tufts a publié des directives sur l'évaluation d'impact participative. Elles soulignent les difficultés d'ordres pratique et éthique à établir des groupes témoins pour tester l'*attribution*, et se concentrent plutôt sur l'utilisation d'outils participatifs pour évaluer la contribution relative de facteurs relatifs ou non à un projet (Catley et coll., 2008).

Encadré 5 Protocoles de recherche pour enquêter sur l'*attribution*

Concept expérimental :

Dans un concept expérimental, les participants se voient assignés de manière aléatoire soit à un groupe qui reçoit les services d'un programme, soit à un groupe témoin qui n'en bénéficie pas et qui sert donc de « condition négative ». On compare ensuite les résultats découlant de ces deux groupes. Cette expérience est conçue de telle manière qu'elle permet d'attribuer les différences de résultat éventuelles entre le groupe récipiendaire et le groupe témoin aux services reçus.

16 Fait intéressant, l'évaluation a permis de conclure que « les principales hypothèses concernant l'utilisation par les réfugiés de l'assistance alimentaire ne se sont pas vérifiées » (Steen et coll., 2013 p. 14).

Modèle quasi expérimental :

Les études quasi expérimentales ont elles aussi pour objet de démontrer l'attribution en comparant les résultats, mais n'impliquent pas l'affectation aléatoire à des groupes. Au lieu de cela, elles comparent les résultats de groupes qui ont reçu des services et de groupes similaires qui n'en ont pas bénéficié (« expérience naturelle ») ; ou encore, les résultats d'un seul groupe avant et après une intervention.

Démarches fondées sur la théorie :

Ces approches ne cherchent pas à démontrer l'attribution en comparant les groupes récipiendaires et les groupes contrefactuels, mais testent plutôt la théorie sous-jacente de causalité en vertu de laquelle les concepteurs de programme s'attendent à ce que certaines activités entraînent certains résultats. Dans une démarche fondée sur la théorie, la série de suppositions au centre de la conception du programme qui relie apport, contexte et résultat est traitée comme hypothèses pouvant être testées à l'aide d'une palette de méthodes quantitatives et qualitatives.

Approches individualisées :

Les approches individualisées reposent sur une étude de ce qui s'est réellement passé dans des cas précis : elles identifient les facteurs qui ont entraîné certains résultats et les comparent ensuite au sein de plusieurs cas ou entre eux afin d'émettre des « généralisations analytiques ».

Sources : définitions à partir de Stern et coll. (2012) ; Morra Imas et Rist (2009), Leeuw (2012).

Toutefois, cette approche participative ne sert pas seulement à contourner des difficultés pratiques. Les évaluations de ce type tiennent compte du fait que l'impact – portant en grande partie sur le plus long terme et sur les conséquences souvent inattendues de toute intervention – ne saurait cadrer parfaitement avec la logique d'un projet élaboré de l'extérieur, et est en général mieux appréhendé par l'expérience des récipiendaires de l'aide eux-mêmes.

Alors que « les systèmes classiques de suivi et d'évaluation ont tendance à surdimensionner 'nos indicateurs' aux dépens de 'leurs indicateurs' » (Catley et coll. s.d. : 21), l'approche participative s'appuie sur des indicateurs d'impact élaborés par les bénéficiaires visés. De même, la People First Impact Method (PFIM) demande aux communautés de définir les changements les plus importants survenus dans leurs vies et leurs causes (voir, par exemple, O'Hagen et McCarthy, 2012). C'est une démarche similaire que la FAO a utilisée en Somalie, où l'évaluation d'impact misait sur une approche itérative et narrative : les récipiendaires de l'aide faisaient part de leur expérience du programme et des incidences que celui-ci avait eues lors d'entretiens semi-structurés complétés de divers outils participatifs. Les thèmes urgents étaient ensuite renvoyés aux groupes communautaires pour être débattus et vérifiés de manière approfondie (Tessitore, 2013).

EN BREF :

- Si une grande partie des orientations relatives à l'évaluation humanitaire soulignent les avantages des approches méthodologiques mixtes, la plupart des évaluations menées ne reposent que sur des méthodes qualitatives.
- Normalement, les évaluations humanitaires utilisent l'échantillonnage au jugé pour assurer leur *représentativité*, et la triangulation pour accroître leur *exactitude*.
- Cependant, bien des évaluations humanitaires ne tirent pas le meilleur parti de ces méthodes qualitatives et, en conséquence, la qualité des données qu'elles soumettent est médiocre.
- Bon nombre d'évaluations ne tiennent pas véritablement compte des perceptions des populations affectées par les crises.
- L'action la plus efficace menée pour structurer l'évaluation autour des perceptions des populations touchées s'observe dans le domaine de l'évaluation d'impact.
- L'*attribution*/contribution présente des défis particuliers pour les évaluations, qui ont pour objet de rechercher les liens de causalité entre un programme et une situation qui s'ensuit.
- On observe depuis peu une montée en puissance des démarches dites expérimentales pour établir l'*attribution*. De plus, les agences commencent à envisager diverses autres approches non expérimentales de l'évaluation de la contribution.



Les données probantes actuelles sont-elles exploitées pour orienter les opérations et l'élaboration des politiques humanitaires ?

4.1 La qualité n'est pas un gage d'utilisation

La section précédente s'est intéressée à la qualité des données générées au sein du système humanitaire international. Si les décisions doivent être informées par des données probantes, il est bien entendu essentiel que ces dernières soient de la plus haute qualité possible. Or, la qualité ne garantit pas à elle seule que ces données soient utilisées. Les processus de prise de décisions sont rarement, voire jamais, tout à fait rationnels : ils sont influencés par diverses considérations organisationnelles et politiques, et par la partialité des décideurs individuels.

Pour que des preuves puissent utilement informer l'action humanitaire, il convient dès lors de s'intéresser non seulement au processus de collecte et d'analyse des informations, mais aussi à celui de prise de décision. L'expérience de la médecine est à cet égard révélatrice : la tendance vers la médecine dite « fondée sur des données probantes » observée dans les années 1990 devait beaucoup aux évolutions dans les procédures organisationnelles et aux changements de mentalité chez les praticiens individuels (Anthony Redmond, communication personnelle, octobre 2012) : « la prise de décision fondée sur des données probantes ne requiert souvent aucune donnée scientifique supplémentaire proprement dite, mais plutôt... de ressources financières et de la volonté politique » (Bradt, 2009 : 3).

Dans cette optique, la présente section examine le degré auquel différents types de données et de savoirs ont été exploités par les décideurs humanitaires et analyse les facteurs qui semblent restreindre, et ceux qui semblent soutenir, l'utilisation de données probantes dans la prise de décision.

4.2 Exploitation des données probantes émanant de l'alerte précoce

Comme on l'a constaté dans la section 3, des progrès considérables ont été réalisés au cours des vingt dernières années en matière de production de données opportunes et exactes pour l'alerte précoce, ce qui a permis de formuler des propositions de type A avec un certain degré de confiance dans de nombreux contextes. Pourtant, l'usage de telles informations s'est avéré plus problématique, notamment dans le cadre de la famine et de l'insécurité alimentaire. Si ce problème de décalage entre les informations issues de l'alerte précoce et les réponses effectives est reconnu depuis longtemps (voir Buchanan-Smith et Davies, 1995), il



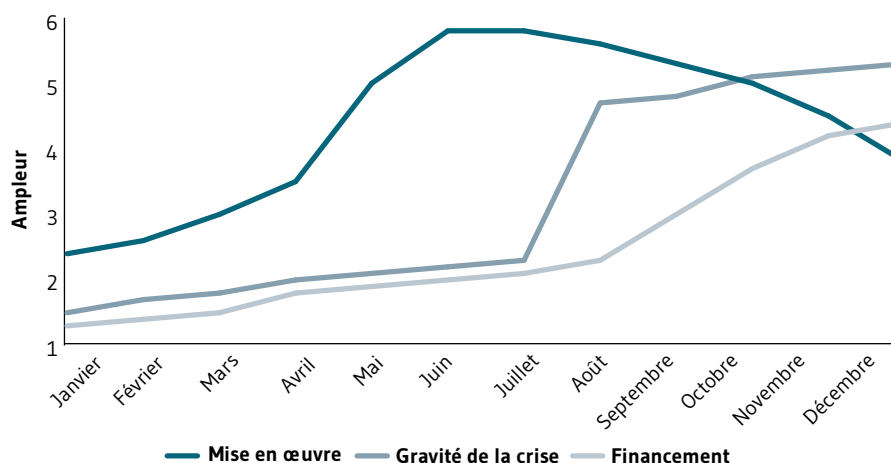
Ce problème de décalage entre les informations issues de l'alerte précoce et les réponses effectives persiste au cœur des politiques et de la prise de décision.



persiste au cœur des politiques et de la prise de décision – par exemple, dans les crises alimentaires à survenue lente (Levine et coll., 2011).

On a pu l'observer dans le cadre de la famine de 2011 dans le centre-sud de la Somalie. La figure 1, extraite de l'évaluation en temps réel (ETR) de la réponse à la crise somalienne menée par l'IASC, présente le lien entre la gravité de la crise, le financement de l'intervention et la mise en œuvre de celle-ci.

Figure 1 : Représentation sous forme de schéma de la réponse à la famine en Somalie de 2011



Source : IASC Évaluation en temps réel de la réponse internationale à la crise de 2011 en Somalie – Valid International, 2012.

Ce qui frappe le plus dans la figure 1, ce sont les écarts entre, d'une part, la gravité de la crise¹⁷ (ligne rouge) et la disponibilité des fonds (ligne verte) et, d'autre part, entre le financement et la mise en œuvre de la réponse (ligne bleue). Si des données solides (en provenance de FEWS Net et de l'IPC de la FAO) sur l'imminence de la crise étaient disponibles grâce à l'alerte précoce dès le dernier trimestre de 2010, la forte augmentation du financement n'est arrivée qu'au moment de la déclaration de famine en juillet 2011. Ce retard a constitué un facteur primordial dans ce que les auteurs de l'évaluation en temps réel ont qualifié « d'échec systémique de la réponse initiale », et ce à deux niveaux :

- échec de l'action préventive visant à s'attaquer aux causes immédiates de vulnérabilité via une intervention d'urgence au niveau des moyens d'existence qui renforcerait la résilience à court terme et réduirait le besoin de secours ;
- échec des premiers secours d'urgence intensifiés à combattre les symptômes les plus aigus de la crise au moment où une telle assistance était la plus nécessaire du début au milieu de 2011 (Valid International, 2012 : 37).

En fait, la planification d'une intervention massivement amplifiée n'a eu lieu qu'en juillet 2011, après la phase la plus aiguë. Les auteurs de l'ETR font la constatation suivante : « Au regard des décalages impliqués, il aurait fallu réaliser celle-ci en janvier/février au plus tard afin d'aborder les agendas de prévention et de secours

17 Cette « gravité de la crise » est fondée sur un panier d'indicateurs comprenant, notamment, la malnutrition aiguë et le prix des denrées alimentaires sur les marchés. Ces trois variables sont reconnues comme étant à la fois difficiles à quantifier et incommensurables, aussi la figure 1 n'est-elle fournie qu'à titre indicatif.

au moment requis. » Quand les programmes d'intervention ont été amplifiés, la crise avait déjà bien dépassé son paroxysme.

La situation au sud-centre de la Somalie en 2010-2011 n'a pas été facilitée par des facteurs politiques et sécuritaires. De grandes parties de la région sinistrée étaient à ce moment-là contrôlées par le groupe Al Shabbaab. Certains financeurs majeurs ont refusé d'autoriser l'aide à parvenir à Al Shabbaab. Le groupe islamiste, lui, a interdit à de nombreuses agences d'opérer dans la région. Pourtant, ce schéma d'alerte précoce et d'intervention tardive a pu s'observer auparavant, dans des contextes où de telles contraintes n'avaient pas la même ampleur (Hillier et Dempsey, 2012) et il s'est reproduit dans d'autres régions de la Corne d'Afrique. Autrement dit, les circonstances propres à la Somalie n'expliquent pas ce phénomène.

Il ne s'agissait pas non plus, pour l'essentiel, d'un problème de disponibilité ou de qualité des données probantes. Comme nous l'avons dit, les données émanant de l'alerte précoce étaient disponibles en temps utile et perçues, dans l'ensemble, comme de bonne qualité – bien que la validité de certaines données et analyses ait suscité des interrogations (Levine et coll., 2011). « En définitive, les systèmes d'alerte précoce ont bien fonctionné, mais les décideurs ont choisi de ne pas y répondre » (Hillier et Dempsey, 2012 : 14).

La non-exploitation par le système humanitaire international des données issues de l'alerte précoce aurait plusieurs causes. Ces contraintes sont liées à la nature des données probantes disponibles, aux processus de prise de décision répandus dans le système humanitaire international, et aux processus opérationnels plus larges du système global.

La première contrainte – la nature des données disponibles – concerne le fait que, même si les organisations savaient que la situation allait en toute probabilité avoir des incidences considérables et négatives sur les vies et les moyens de subsistance des populations de la région, elles ne savaient pas quoi faire de cette information, c'est-à-dire comment traduire ces savoirs en interventions efficaces (Levine et coll., 2011). En effet, s'il existe des exemples documentés d'action préventive dans des situations de ce type (voir la discussion sur l'intervention précoce dans les situations d'urgence liées à la sécheresse à la section 4.4), il semble que la plupart des organisations n'aient pas été conscientes de ces précédents ou n'aient pas eu l'habitude de les utiliser, ayant traditionnellement privilégié l'intervention après la survenue d'une crise. D'où la difficulté qu'éprouvent bien des organisations humanitaires à muer les informations obtenues de l'alerte précoce en programmes concrets.

Au vu des propositions humanitaires énoncées précédemment, ce constat est important : dans la Corne d'Afrique, il n'était pas possible d'exploiter les données probantes relatives à la proposition de type A (selon laquelle un problème existe) sans avoir aussi des données probantes relatives aux propositions de type B (selon lesquelles une certaine approche est efficace pour lutter contre le problème). L'alerte précoce aurait peut-être été plus efficace si elle avait été assortie d'exemples de bonne pratique programmatique obtenus auprès des agences ayant de l'expérience dans ce domaine.

Cela n'était cependant pas la seule cause d'inaction. Le manque de réponse en Somalie met aussi en évidence la seconde contrainte : des faiblesses au niveau de la façon

dont les décisions sont prises dans le système humanitaire. Selon Hillier et Dempsey, « beaucoup de gens 'sur le terrain', dans la région... étaient conscients de la crise imminente et ont essayé de tirer la sonnette d'alarme en janvier et février 2011, [mais] n'ont pas toujours pu susciter de l'intérêt 'un peu plus en amont' » (Hillier et Dempsey, 2012 : 4). En raison de la nature mondiale et atomisée du système humanitaire, les décisions requièrent souvent l'aval d'un grand nombre d'acteurs situés dans des organisations et des lieux différents ; et savoir par qui et où doivent être prises les décisions n'est pas toujours clair (Knox Clarke, 2013b). De plus, les individus qui occupent des postes différents ne sont pas motivés dans leur tâche par les mêmes incitations et sont susceptibles d'accorder des poids différents à une même donnée.

En somme, dans le cas somalien, il semble que ce qui a incité le personnel sur place à tenir compte des données issues de l'alerte précoce, c'est la détérioration de la situation dont il était témoin sur le terrain. Les membres de l'équipe humanitaire pays (EHP), en revanche, étaient conscients de la difficulté à lever des fonds et étaient plus réticents à exploiter les informations (Slim, 2012).

Certains décideurs en Éthiopie ont estimé que les besoins détectés par les systèmes d'alerte précoce étaient « trop petits » pour catalyser une réaction efficace parmi les bailleurs (Majid, 2011). En même temps, soucieux du besoin de tirer le meilleur parti de l'argent du contribuable, les bailleurs des capitales occidentales étaient réticents à débloquent des fonds tant qu'ils n'avaient aucune preuve (sous forme de données nutritionnelles) qu'une crise humanitaire avait déjà lieu (Hillier et Dempsey, 2012 ; Levine et coll., 2011). De même, les gouvernements de la région étaient peu enclins à faire appel à l'assistance internationale à moins que cela soit absolument nécessaire, de peur qu'un tel appel soit perçu comme une admission d'échec de leur part.



Le niveau élevé d'évitement des risques dans le système humanitaire, tend à transférer le risque des organisations humanitaires aux populations touchées par des catastrophes.



À tous les niveaux, les acteurs devaient décider quelle attention accorder à des données précises dans une situation d'incertitude, et leurs décisions étaient fortement influencées par d'autres informations et des motivations organisationnelles. Il s'agit d'un problème commun dans les opérations humanitaires : de nombreux auteurs ont en effet évoqué le niveau élevé d'évitement des risques dans le système humanitaire (se reporter à Buchanan-Smith et Scriven, 2011), comportement qui, on pourrait le soutenir, tend à transférer le risque des organisations humanitaires aux populations touchées par des catastrophes.

En 2010-2011, ces problèmes de prise de décision furent exacerbés par le manque de « déclencheurs » (Levine et coll., 2011), c'est-à-dire de déclarations claires de l'ampleur des besoins, notamment dans des secteurs comme l'eau et l'assainissement, qui auraient automatiquement déclenché une décision sur l'intervention. De fait, les systèmes d'alerte précoce faisaient partie d'un processus de prise de décision plus global : ils semblent hélas avoir été considérablement mieux développés que d'autres éléments de ce processus. Un manque particulier de clarté entourait certaines questions clés : qui doit prendre les décisions, l'importance relative de différents critères d'évaluation des données afin de prendre des décisions, et même les questions de savoir quand prendre les décisions, voire s'il fallait les prendre.

Le troisième groupe de contraintes pesant sur l'utilisation des données probantes émanant de l'alerte précoce avait trait aux processus opérationnels plus larges qui convertissent des fonds en mesures concrètes. En général, le déblocage de fonds et le lancement d'activités sont actuellement subordonnés à une évaluation des

besoins plutôt qu'à des mécanismes d'alerte précoce (Levine et coll., 2011). Dès lors, on constate normalement un intervalle de plusieurs mois avant que les opérations démarrent sur le terrain. Vu sous cet angle, les données probantes fournies par l'alerte précoce ne servent pas à prendre des décisions concernant une intervention, parce que l'alerte précoce n'est pas conçue formellement pour aboutir à ces décisions.

Pour remédier à cela, Levine et ses collègues proposent une série de mesures concrètes de préparation dans des domaines tels que le recrutement et la passation des marchés, mesures que les agences pourraient prendre sur la base des données fournies par l'alerte précoce afin de réduire leur temps de réaction. De même, Hillier et Dempsey préconisent une série de mesures « Sans regret », notamment de renforcement des capacités et de prépositionnement des stocks, qui devraient être déclenchées par l'alerte précoce.

EN BREF :

- Bien que la qualité des systèmes d'alerte précoce se soit améliorée, les données probantes signalant la survenue vraisemblable d'une crise se traduisent rarement en action, comme l'a attesté le manque de réaction à de telles données dans la Corne d'Afrique, et surtout dans le sud-centre de la Somalie, en 2010-2011.
- La non-exploitation des données fournies par l'alerte précoce a des causes diverses, mais l'un des facteurs clés est que la production de données probantes à partir de l'alerte précoce ne s'inscrit pas dans un processus plus large et structuré d'alerte précoce et d'intervention :
 - le « processus administratif » actuel de l'intervention humanitaire couple l'intervention à des activités d'évaluation et non à l'alerte précoce et le rôle des données issues de cette dernière dans le déclenchement d'une intervention n'est pas encore clair
 - L'identité du décideur est rarement claire en ce qui concerne la réponse initiale, ainsi personne n'est clairement responsable de ces décisions
 - Les mesures à prendre à partir de l'alerte précoce sont elles aussi rarement claires : les organisations humanitaires ont plus d'expérience en matière de réponse aux crises qu'en matière de prévision.
- Enfin, outre des échecs structurels et procéduraux, les motivations organisationnelles (et en particulier le souci d'éviter le risque de « gaspiller de l'argent ») découragent les actions éventuelles à mener avant la survenue d'une crise.

4.3 Exploitation des données probantes fournies par l'évaluation et le suivi

Si la place de l'alerte précoce dans le processus de prise de décision est souvent floue, aucune confusion n'entoure en revanche le rôle des évaluations des besoins, conçues précisément pour fournir des données probantes utiles à la prise de décision. De manière générale, ces données portent sur les propositions de type A relatives à l'existence de besoins, et permettent de décider si une assistance est nécessaire, quelle doit être son ampleur¹⁸, à qui elle doit être destinée, et pour

18 Bien que, comme nous l'avons vu, l'erreur commune des évaluations est de supposer qu'une assistance est requise et de se polariser plus sur l'ampleur de l'assistance à fournir.

combien de temps. L'évaluation est souvent prise en compte par au moins deux groupes de décideurs : le personnel d'agence, qui y fait appel pour concevoir le programme initial et pour demander des fonds ; les bailleurs ensuite, qui l'utilisent afin de déterminer si les demandes de fonds sont légitimes. De plus, les praticiens humanitaires comptent souvent pouvoir utiliser les informations découlant des évaluations des besoins comme preuves leur permettant d'étayer les décisions portant sur les propositions de type B : le meilleur type d'assistance à apporter dans le contexte spécifique donné. Ce dernier usage est néanmoins problématique, et nous y reviendrons plus loin.

Une analyse documentaire permet de penser que, s'agissant des décisions relatives aux propositions de type A (y aller ou pas, où, et ampleur de l'intervention) l'usage des évaluations des besoins est variable, et que lorsqu'il est fait appel à des évaluations, les données qu'elles fournissent ne représentent souvent qu'un facteur parmi plusieurs qui influencent les décisions. Ceci pourrait surprendre à premier abord – après tout, les évaluations sont principalement conçues comme des outils de prise de décision –, mais s'explique par diverses contraintes liées tant aux évaluations mêmes qu'à la nature des décisions humanitaires.

Comme on l'a vu, la qualité de preuve des évaluations des besoins humanitaires est inégale et, puisqu'elles ne sont pas souvent menées par des acteurs impartiaux, mais par les mêmes agences qui les utiliseront par la suite pour justifier leurs demandes de financement, les évaluations sont souvent traitées avec scepticisme par les bailleurs (Bradt, 2009 ; Darcy et coll., 2013). De ce point de vue, il ne fait aucun doute que s'attaquer aux problèmes liés à la qualité des données probantes exposés à la section 3.1 permettrait d'améliorer le recours aux évaluations dans la prise de décision.

Seconde contrainte inhérente au processus d'évaluation même (plutôt qu'à des facteurs largement indépendants de l'évaluateur) : l'obtention des données en temps utile. En effet, pour que les évaluations influencent la prise de décision, l'information doit être disponible lorsque la décision est prise.

Dressant le bilan du recours à une évaluation initiale rapide des besoins conduite par ACAPS à la suite du tremblement de terre de 2010 en Haïti, Darcy et Garfield observent l'effet des retards à mettre les résultats de l'évaluation à la disposition des clusters et des autres décideurs. Ils concluent que « l'on ne sait pas très bien si l'analyse aurait informé la prise de décisions portant sur l'Appel éclair ou les plans de cluster si elle avait été disponible plus tôt. Ce qui est certain c'est que, même en admettant que l'analyse ait été pertinente et crédible, elle est arrivée trop tard pour informer les décisions de planification initiales » (Darcy et Garfield, 2011).

Une démarche progressive pour l'évaluation, telle celle décrite dans les orientations du MIRA (IASC, 2012a) peut contribuer à éliminer cette contrainte, en assurant que certaines informations sont disponibles dans les 72 heures, des analyses de plus en plus poussées étant fournies dans les semaines et mois qui suivent. Bien entendu, les informations qui peuvent être fournies dans les 72 premières heures sont assez limitées : les évaluations des besoins ressentis lors de crises à survenue rapide, peut-être plus que dans tout autre domaine de l'effort humanitaire, mettent les praticiens au défi de trouver un équilibre entre la qualité et la rapidité – l'objectif devant être de trouver des données probantes qui « suffisent » pour étayer la prise de décision (Lars Peter Nissen, communication personnelle, décembre 2012).

Dans bien des contextes humanitaires, le temps est précieux. Non seulement il décourage les décideurs à attendre des preuves, mais il les dissuade aussi de lire de longs rapports d'évaluation. Si des données probantes doivent être utilisées, alors leur présentation est importante. ACAPS a constaté que les rapports courts qui présentaient les informations sous forme graphique avaient plus de chances d'influencer la prise de décision (ACAPS, 2012). Une étude de l'Humanitarian Policy Group (HPG) sur les diagnostics humanitaires, fait observer qu'en général « la manière dont l'information est présentée est parfois déterminante pour son degré d'utilisation par les décideurs », et préconise des évaluations courtes, succinctes contenant des « faits chocs » pour attirer l'attention (Darcy, 2009 : 11).

Dans certains cas, notamment lorsque le décideur se trouve au sein de l'organisation qui a procédé à l'évaluation, ou lorsqu'il est dans l'ensemble prédisposé à exploiter les informations issues de l'évaluation, il suffit de veiller à ce que les rapports soient solides, clairs et opportuns pour assurer que les données qu'ils contiennent sont utilisées. Or, dans la plupart des cas, ce degré d'utilisation est également déterminé par des facteurs liés davantage au décideur – et au contexte dans lequel celui-ci prend ses décisions – qu'à l'évaluation elle-même. En particulier, les données issues de l'évaluation ont plus de chance d'être exploitées lorsqu'elles recommandent des mesures possibles ou souhaitables pour les décideurs et lorsqu'elles correspondent à leurs suppositions (souvent inconscientes) sur la situation.

Plusieurs auteurs ont récemment relevé l'importance, dans la prise de décision, de la notion de dépendance au chemin emprunté pour déterminer le degré de prise en compte des informations émanant d'évaluations (Darcy et coll., 2013 ; Maxwell et coll., 2013). Pour l'essentiel, cette notion reconnaît que les décisions humanitaires ne sont pas prises dans le vide, mais dans un contexte où certains éléments sont fixés au préalable (ou du moins difficiles à modifier). La plupart des décisions sont fortement influencées par des décisions qui ont déjà été prises en matière de stratégie organisationnelle, de capacité ou d'allocations de fonds.

Lorsque l'on délibère sur des questions comme l'échelle, le lieu et la nature des interventions, « la plupart des décisions [prises au sein d'agences] le sont dans des paramètres très serrés, la panoplie d'options étant limitée par des questions décidées antérieurement portant, entre autres, sur les priorités stratégiques et les ressources disponibles » (Darcy et coll., 2013 : 7). De même, quand elles délibèrent sur le type d'intervention à mettre sur pied, les agences ne partent pas d'une feuille blanche. Au lieu de cela, « les choix de réponse sont également déterminés par la capacité et l'éthique organisationnelles de l'agence de mise en œuvre » (Maxwell et coll., 2013 : 7).

Et même lorsqu'il y a plus de marge de prise de décision – lorsque les décisions antérieures ne déterminent pas l'allocation des ressources – une évaluation des besoins n'est « que le début d'une série d'informations, influences et considérations que... [les décideurs] doivent apprécier » (Poole et Primrose, 2010 : 5). Poole et Primrose observent que, pour allouer des fonds, les financeurs doivent tenir compte de divers facteurs, lesquels peuvent avoir une influence légitime sur leurs décisions. Même si les évaluations indiquent un besoin d'intervention dans un certain lieu et à une certaine échelle, les bailleurs peuvent choisir de ne pas le faire parce que les besoins sont plus grands ailleurs, ou parce que les conditions de sécurité ou le manque de partenaires fiables rendent peu probable la réussite d'une intervention. Des considérations politiques et médiatiques peuvent également entrer en jeu dans la décision.

Ainsi, la prise de décision humanitaire requiert-elle en réalité l'appréciation consciente de divers ensembles de données – en l'occurrence des données sur les besoins, les capacités et les conditions sécuritaires – avant que tout jugement soit prononcé. Dès lors, il convient de concevoir les bonnes décisions comme étant « informées par des données probantes » plutôt que « reposant sur des données probantes », car différentes données ont tendance à étayer des propositions concurrentes, même si elles sont toutes de véracité égale. Cela étant, un éventail d'éléments moins rationnels et souvent moins conscients influence toutefois fréquemment la prise de décision.



Il convient de concevoir les bonnes décisions comme étant « informées par des données probantes » plutôt que « reposant sur des données probantes », car différentes données ont tendance à étayer des propositions concurrentes, même si elles sont toutes de véracité égale.



Travaillant contre la montre et dans des situations où ils doivent examiner une grande quantité d'informations, les décideurs ont tendance à prendre des raccourcis – simples hypothèses, généralisations et règles empiriques qui permettent d'organiser les informations et de leur donner un sens : ce sont les heuristiques, terme employé par les psychologues. Ces heuristiques servent de filtre et ont comme effet général de diminuer la capacité des personnes à remarquer des informations imprévues, insolites ou problématiques et à s'y intéresser (Fiske et Taylor, 1991 ; Macrae et coll., 1993), surtout lorsque des décisions sont requises et que le temps est limité (Bohner et coll., 1995). La tendance humaine à suivre des « modèles mentaux » existants – même s'ils sont remis en question par de nouvelles données – a pu s'observer dans la sphère humanitaire (Darcy, 2009 ; Maxwell et coll., 2013 ; Knox Clarke, 2013b), et semblerait influencer la capacité à absorber les données issues des évaluations de trois manières précises.

La première concerne ce que Peter Walker nomme les « modèles de départ ». Nous avons déjà constaté que les évaluations des besoins se fondent souvent sur une série d'hypothèses – la plus importante étant l'hypothèse souvent incorrecte qu'il n'existe aucune capacité locale de réaction à la crise. Cela est vrai aussi pour les opérations humanitaires en général : elles commencent par des suppositions – ou modèles de départ – sur une situation, lequel devient un type d'heuristique : une façon d'ordonner l'information. Dans ce contexte, les données découlant d'évaluations des besoins qui vont à l'encontre du modèle de départ tendent à être ignorées. Un exemple : la situation en Afghanistan, où, après l'intervention militaire en 2001, « le modèle était que le conflit était terminé, que l'on pouvait travailler avec le gouvernement, qui n'est pas un combattant » (Walker, 2013).

Les modèles de départ ont la vie d'autant plus dure qu'ils ont tendance à être partagés par un grand nombre de personnes dans le secteur humanitaire. D'où une seconde heuristique : celle des connaissances sociales. Lorsqu'un scénario a été largement admis, toute information qui le réfute aura du mal à s'imposer. Poole et Primrose citent « la conception très répandue du scénario de la transition rapide vers le développement » (Poole et Primrose, 2010 : 1) au Soudan du Sud, qui a rendu difficile la mise en relief des besoins humanitaires. On peut observer une dynamique similaire en République démocratique du Congo (RDC), où un « scénario post-conflit » a entraîné une situation où « le Plan d'action humanitaire en RDC manque désormais régulièrement de fonds, alors que les besoins humanitaires demeurent aussi élevés que jamais » (Darcy et coll., 2013 : 10).

Dernière heuristique, et peut-être la plus répandue, celle de « l'ennemi familier », selon laquelle les décideurs ont tendance à interpréter les situations à l'aune

d'expériences passées. Ces expériences constituent en elles-mêmes un type de données probantes – mais ces données ne sont pas toujours pertinentes par rapport à la nouvelle situation, et elles ne sont pas toujours solides. De plus, les décideurs ont tendance à se rabattre sur des méthodes éprouvées pour réagir aux circonstances (Maxwell et coll., 2013) et, lorsque ce n'est pas le cas, exigent des niveaux de preuves plus élevés avant d'approuver de nouvelles marches à suivre (Darcy, 2009). Cette tendance à ne pas faire cas d'informations qui laissent entrevoir des conditions ou des mesures inhabituelles peut, bien entendu, être entretenue par les mesures qui incitent, au niveau de l'organisation, à éviter la prise de risque..

Tous les facteurs cités plus haut – dépendance au chemin emprunté, priorités concurrentes et diverses heuristiques – s'appliquent aux décisions liées aux propositions de type A (une intervention est requise ou non) et de type B (quel type d'intervention mener). Le recours aux évaluations est cependant également soumis à une contrainte qui est spécifique aux propositions de type B. Cette contrainte est au cœur des travaux de Maxwell sur l'analyse de l'intervention : un décalage entre « le type d'informations normalement fournies par les évaluations et le type d'informations que le choix d'intervention nécessite » (Maxwell et coll., 2013 : 1).

Il ne s'agit pas d'un problème de qualité des données, et ce n'est pas que les données ne peuvent pas donner lieu à des mesures ou qu'elles sont ignorées. C'est plutôt que les données probantes requises pour agir sont incomplètes. Le processus d'évaluation des besoins ne permet pas en effet de recueillir les données dont ont besoin les décideurs, tout comme les systèmes d'alerte précoce ne peuvent pas fournir d'informations sur ce que les organisations humanitaires devraient faire pour empêcher une crise. Dans une certaine mesure, cette carence est le reflet des silos qui existent entre les différents spécialistes de l'information au sein du système humanitaire : les experts de l'évaluation des besoins se concentrent étroitement sur les besoins, laissant le soin d'une intervention au personnel du programme, de même que les évaluateurs travaillent souvent séparément de ceux qui gèrent les programmes (Argren, 2013 ; Hallam et Bonino, 2013).

Quelle qu'en soit la cause, les décideurs n'obtiennent pas de ce fait toutes les informations dont ils ont besoin. L'importance du recueil de données probantes, pas seulement pour étayer des propositions générales, mais pour aborder les questions précises qui préoccupent les décideurs, est un thème que nous continuerons d'examiner dans la prochaine section.

EN BREF :

- L'utilisation des évaluations des besoins dans la prise de décision est variable.
- Cette variabilité est due en partie aux problèmes inhérents aux évaluations elles-mêmes : elles ne sont pas toujours perçues comme impartiales ; elles ne sont pas toujours soumises en temps voulu ; elles sont parfois inaccessibles et de lecture ardue ; enfin, elles ne contiennent pas toujours les informations dont le personnel des agences a besoin pour planifier les programmes.
- La variabilité résulte également de la façon dont les décisions sont prises :
 - les décideurs ont souvent une marge de manœuvre limitée en ce qui concerne la prise de décision, de nombreux choix ayant déjà été exclus par les stratégies ou les missions organisationnelles.

- Lorsqu'ils prennent des décisions, ils doivent aussi prendre en compte une gamme d'informations – les données sur les besoins ne constituent qu'une partie d'un ensemble plus large d'informations.
- Les décideurs sont influencés par des préjugés inconscients qui peuvent les amener à négliger certaines informations issues des évaluations.

4.4 Utilisation des données issues des évaluations

Le nombre d'évaluations humanitaires s'est considérablement accru au cours des dix dernières années. La bibliothèque des ressources d'ALNAP – en aucun cas exhaustive – abrite plus de 1200 évaluations de l'action humanitaire. À ce titre, elle représente l'une des sources les plus importantes de données sur les pratiques efficaces (et inefficaces) dans la réponse humanitaire internationale.

Comme l'ont observé Telford et Cosgrave dans le rapport de synthèse de la Tsunami Evaluation Coalition (TEC), dans un contexte où beaucoup de rapports d'agences « se concentrent sur les succès et ne font pas cas des échecs ou les passent sous silence... [...] et la couverture médiatique a tendance à privilégier les cas uniques spectaculaires plutôt qu'un bilan équilibré de la qualité globale, les informations les plus détaillées sur la performance d'une agence pourraient bien provenir des rapports d'évaluation d'agences (Telford et Cosgrave : Synthèse 2006 : 108).

Si le nombre d'évaluations est en hausse, un certain scepticisme semble planer sur le degré réel de leur utilisation. L'usage des données émanant d'évaluations touche à la fois les pratiques opérationnelles et l'élaboration des politiques dans le secteur humanitaire. Sandison (2007), emboitant le pas à Patton (1997), décrit trois principaux usages des données de l'évaluation :

- i. juger le mérite ou la valeur d'un programme (redevabilité envers les parties prenantes ;
- i. améliorer un programme (par exemple, apprentissage et formation continue) ;
- ii. générer des savoirs (pouvant être utilisés dans des programmes ultérieurs et ailleurs).

Comme le note Sandison, les usages (i) et (ii) peuvent être considérés comme « instrumentaux » : l'évaluation sert à modifier directement les programmes, secteurs ou systèmes qui font l'objet d'une évaluation (modifications souvent fondées sur les recommandations des évaluateurs). La plupart des évaluations dans le secteur humanitaire sont conçues principalement pour ce type d'usage instrumental.

Le troisième type d'usage (iii) est, lui, plus général et moins direct. Dans ce cas-là, les informations contenues dans l'évaluation s'ajoutent à un stock général de connaissances sur un sujet ; elles servent à apporter des changements et des améliorations indirectement, de loin. Si ce type d'usage est peut-être considéré comme relevant davantage du domaine de la recherche, les évaluations contribuent souvent beaucoup à la production de savoirs (voir par exemple, les rapports d'ALNAP sur les enseignements tirés : (Hedlund et Knox Clarke, 2011 ; Sanderson et Knox Clarke, 2012).

Dans l'optique de la présente étude, les deux premiers types d'usage auront vraisemblablement un seuil probant plus bas que le troisième. Si les données probantes visant à informer des changements dans un programme ou des pratiques précis doivent satisfaire (dans l'idéal) des standards minimums d'*exactitude*, de *représentativité* et de *pertinence*, et si elles doivent aborder efficacement les questions d'*attribution*, elles n'ont pas besoin d'être généralisables. La troisième catégorie est différente, et la mesure dans laquelle elle est perçue comme contribuant aux connaissances au-delà de la situation précise évaluée dépend du degré de satisfaction de tous les critères – y compris celui de *généralisabilité* – de qualité probante.

L'usage instrumental des évaluations

Concernant l'utilisation instrumentale des évaluations, les conclusions de Sandison sont assez négatives : « Seule une minorité d'évaluations parvient à apporter des modifications ou des améliorations aux performances » (Sandison, 2006 : 91). Et d'ajouter : « l'usage instrumental est la forme la moins probable d'utilisation » (ibid : 121). Toutefois, l'auteure constate aussi que « nous ne savons même pas combien d'évaluations sont menées, et encore moins combien sont utilisées » ; dès lors, « la source de préoccupation concernant leur non-utilisation dans le secteur est pour l'essentiel empirique » (ibid : 91).

D'ailleurs, s'il n'est pas difficile de trouver des exemples de recommandations émanant d'évaluations qui n'ont pas été prises en compte, et si « en général les écrits décrivent un usage variable, voire décourageant, de l'usage des évaluations » (Sandison, 2006), le tableau n'est en aucun cas entièrement négatif. Les évaluateurs professionnels que nous avons contactés au cours de la préparation de cette étude ont régulièrement fourni des exemples de recommandations qui avaient été suivies. Le rapport de la TEC, par exemple, aurait donné lieu à des améliorations de la capacité d'intensification à travers l'ensemble du système. Il aurait en outre imprimé de l'élan aux travaux du Groupe de travail sur les évaluations des besoins. De même, la seconde évaluation des clusters a permis d'accorder plus d'importance à la coopération avec les autorités locales dans les interventions internationales. Enfin, les évaluateurs qui ont participé à l'évaluation de la communauté de praticiens d'ALNAP et contribué à l'Étude d'ALNAP *Using Evaluation for a Change (Utiliser l'évaluation en faveur du changement*, Hallam et Bonino, 2013) ont pu faire état de nombreuses situations où les évaluations ont été utilisées, et ils ont relevé plusieurs conditions qui semblaient favoriser leur usage.

Le tableau est similaire si l'on considère les relevés quantitatifs, relativement rares, de la mise en œuvre de recommandations issues d'évaluations. La matrice de réponse au bilan intermédiaire du Fonds central d'intervention d'urgence (CERF) d'OCHA a montré que, dans l'année qui a suivi ce bilan, 50 % des recommandations ont été mises en œuvre (OCHA, 2007). Lorsque le PAM a étudié le degré d'adoption des recommandations des évaluations, il a constaté que 54 % d'entre elles avaient été mises en œuvre et 65 % avaient été intégrées dans les nouveaux documents. Le PAM a noté que les recommandations avaient plus de chances d'être suivies si elles étaient opérationnelles, plutôt que stratégiques, et si leur mise en œuvre ne nécessitait une action que d'un nombre limité de personnes. Les recommandations plus larges, ou celles qui demandaient une coordination avec les partenaires ou les unités des sièges, avaient moins de chances d'être suivies, de même que les recommandations sans avantages tangibles ou celles qui impliquaient une critique du personnel du PAM (PAM, 2005).

Ces observations permettent de penser que les données fournies par les évaluations humanitaires servent souvent à effectuer des changements « instrumentaux » au niveau des financements ou des programmes, mais de manière hautement sélective. Pour déterminer si une évaluation est utilisée, il convient peut-être d'accorder moins d'importance à la qualité des données qu'au degré de facilité de mise en œuvre de ses recommandations éventuelles.



Les données fournies par les évaluations humanitaires servent souvent à effectuer des changements "instrumentaux" au niveau des financements ou des programmes, mais de manière hautement sélective.



Au cours des dix dernières années, maintes organisations humanitaires – dont le DFID, le SIDA, l'UNICEF et le PAM – ont tenté d'identifier des moyens d'améliorer le recours aux données d'évaluation et leur usage. ALNAP a par ailleurs publié quatre études sur ce sujet – Hallam et Bonino (2013), Hallam (2011), Sandison (2005), et Van de Putte (2001) – qui reposent sur ces expériences et celles d'autres membres du Réseau. Ces travaux tendent à indiquer que les évaluations ont plus de chances de mener à des changements au niveau de la mise en œuvre ou du financement de programmes dans les cas de figure suivants : lorsqu'un intérêt se manifeste déjà envers la performance d'un programme ou un débat en la matière est déjà engagé ; lorsque la production de l'évaluation coïncide avec un « créneau » de prise de décision (extension d'un programme, par exemple) ; lorsque les résultats sont communiqués dans un format approprié et accessible aux décideurs ; enfin, lorsque des mécanismes de suivi sont en place.

L'enseignement le plus important qui ressort de ces études, cependant, pourrait être qu'il est capital d'impliquer les décideurs opérationnels à chaque étape du processus d'évaluation : de la sélection des questions d'évaluation à la mise en œuvre et au suivi, en passant par la collecte d'informations. Cette implication permet d'assurer la pertinence de l'évaluation par rapport aux besoins opérationnels, et renforce l'appropriation des données. En même temps, l'implication étroite du personnel du programme soulève des questions autour de l'objectivité des données de l'évaluation et – si l'objectivité est perçue comme un élément important de la rigueur méthodologique – peut susciter des inquiétudes quant à la qualité des données de l'évaluation. Comme le fait remarquer un auteur, « Il existe généralement une tension entre l'indépendance des services d'évaluation et leur succès à mobiliser les usagers des évaluations » (Foresti, 2007).

Les évolutions récentes de l'évaluation humanitaire tendent à incorporer certaines de ces approches, voire toutes, dans un effort pour augmenter les chances que ces évaluations soient utilisées. L'évaluation en temps réel (ETR) suscite un intérêt croissant, car elle permet de produire des informations sur l'évolution d'une opération pouvant être exploitées immédiatement pour « corriger la trajectoire ». En avril 2011, l'IASC a intégré les évaluations en temps réel interagences parmi les composants nécessaires de toutes les situations d'urgence à l'échelle du système (niveau 3). Dans ce cas-là, les informations ne devaient pas seulement être disponibles en temps utile, mais les travaux d'évaluation étaient explicitement liés aux prises de décisions, car l'ETR était conçue pour « informer la réunion des responsables au terme de la période d'activation de trois mois » (IASC, 2011). L'IASC souhaite dorénavant progresser vers la mise en œuvre d'examen opérationnels en temps réel, qui seront avant tout lancés par les équipes humanitaires pays, initiative qui pourrait être destinée à accroître l'appropriation des résultats par les pays.

En Haïti, le groupe URD a mis en place ce que Grünewald nomme « l'évaluation en temps réel plus » : une équipe effectue une série d'évaluations du même projet

sur une période de deux ans, en étroite collaboration avec l'équipe du projet au fil du processus. Les évaluations menées plus tard s'attachent surtout à décrire les avancées réalisées par rapport aux recommandations des missions précédentes et de l'identification de nouvelles difficultés. Comme Grünewald l'explique, « entre l'évaluateur et le personnel du programme, s'instaure ainsi un dialogue solide qui se poursuit pendant toute la vie du projet... l'évaluateur perd un peu de son indépendance (mais, de préférence, pas son objectivité) pour devenir un vecteur de changement... les améliorations obtenues – qui sont, après tout, l'objectif principal – justifient cette approche » (Grünewald, 2012, para. 3).

ACF encourage aussi le dialogue entre évaluateurs et personnel sur le terrain, dans le but d'accroître l'utilisation des enseignements tirés des évaluations. Le réseau a modifié son processus d'évaluation afin d'assurer que les évaluateurs identifient régulièrement les pratiques de référence dans le cadre de leur travail. Le personnel responsable du programme est invité à étudier, débattre et développer ces pratiques de référence, lesquelles sont ensuite incorporées à un bilan des enseignements et diffusées sur tout le réseau (voir l'ACF Learning Review, 2011). Ainsi, l'objectivité évaluative est maintenue et le lien vital entre évaluation et apprentissages organisationnels est renforcé considérablement (Guerrero, 2012 ; Allen, 2012).

Données probantes générales et politique – l'exploitation des évaluations et de la recherche

Nous abordons dans cette section les usages non instrumentaux des évaluations, et de la recherche plus généralement. La question déterminante à cet égard est de savoir comment – et dans quelle mesure – les données émanant de l'action humanitaire dans des contextes spécifiques informent la pratique humanitaire à d'autres moments et dans d'autres lieux. Dans quelle mesure, par exemple, des données de bonne qualité de 2000 en provenance d'Afghanistan peuvent-elles influencer la programmation au Zimbabwe en 2012 ?

Les données peuvent informer la pratique par une variété de canaux. Par exemple, les synthèses de données, tels les rapports sur les enseignements tirés mentionnés plus haut, peuvent être lus par les individus sur le terrain et informer directement leurs actions. Les données issues des évaluations peuvent servir à créer des standards, qui permettent par la suite d'orienter l'action en matière de conception et de mise en œuvre de programmes. Dans certains cas, les données peuvent être exploitées pour élaborer des mécanismes ou des logiciels d'aide à la prise de décision et auront par conséquent une influence directe sur les décisions prises à partir de ces outils.

Toutefois l'élaboration de politiques est peut-être la voie la plus importante pour assurer que les données sont mises en pratique : les données probantes informent la politique menée, qui à son tour oriente la pratique dans toute l'organisation. Dans cette section, nous nous penchons sur les façons dont les données évaluatives et autres ont influencé l'élaboration des politiques, en examinant tout particulièrement quatre exemples : la programmation faisant appel à des transferts monétaires, l'adoption de standards minimums, la prise en compte de facteurs culturels dans la programmation humanitaire, et l'intervention précoce dans les situations d'urgence liées à la sécheresse.

Programmation faisant appel à des transferts monétaires

L'une des évolutions les plus significatives des dix dernières années en matière de politique a été l'acceptation et le soutien accrus en faveur du recours aux transferts d'argent à la place des distributions d'aliments et d'autres biens. Par exemple, la politique d'assistance humanitaire d'ECHO (direction générale de l'aide humanitaire et de la protection civile de la Commission européenne), ainsi que les directives de l'organisation sur l'utilisation d'argent, correspondent à une acceptation de plus en plus large de la programmation de transfert monétaire. L'agence a récemment relevé le plafond fixé à 100 000 euros des programmes de transfert d'argent (DG ECHO, 2009). Au Royaume-Uni, le bilan de l'intervention humanitaire d'urgence a recommandé que le DFID « fasse des interventions monétaires la position de secours et de relèvement habituelle pour ses partenaires » (Ashdown, 2011 : 24). USAID a récemment modifié la politique portant sur le titre II de son programme Food for Peace afin d'y inclure expressément la programmation faisant appel aux transferts monétaires.

Un tel soutien politique au niveau des bailleurs a entraîné une hausse marquée du financement de la programmation monétaire dans les opérations humanitaires : selon Development Initiatives, les dépenses dans les programmes de transferts monétaires sont passées de 74,9 millions de dollars en 2006 à 188,2 millions de dollars en 2010 (Development Initiatives, 2012).

Bien que les fonds aient affiché une chute en 2011, la tendance générale semblerait être à un recours accru à l'argent, en particulier parce que les grandes agences planifient de développer considérablement leurs activités dans ce domaine (Haley Bowcock, communication personnelle, décembre 2012). En tant que stratégie de programmation, « l'action monétaire dans les secours humanitaires est passée... de radicale et risquée à... courante » (Ramalingam et coll., 2009 : 43).

Jusqu'à quel point peut-on mettre cette réorientation des politiques sur le compte des données probantes ? Les points clés de l'étude d'ALNAP qui documente l'évolution du recours à l'argent liquide dans la programmation humanitaire méritent d'être répétés ici (Ramalingam et coll., 2009). L'étude affirmait que si l'argent était utilisé depuis assez longtemps dans l'intervention d'urgence, il a fallu attendre l'an 2000 pour que ces expériences atomisées soient étudiées méthodiquement dans un document unique : *Buying Power : the use of cash transfers in emergencies* (Peppiatt et coll., 2000). Les travaux de l'Humanitarian Policy Group (HPG) de l'ODI ont suivi, avec la publication d'une série d'études examinant l'utilité de l'argent dans les contextes d'urgence. Nombreux parmi ceux qui ont participé à l'adoption de la programmation monétaire « ont cité les travaux de... HPG comme étant capitaux... pour persuader plusieurs agences à lancer leurs propres projets » ; de plus, « des recherches crédibles documentant la viabilité de l'argent dans divers milieux... ont aidé les organisations dans leur action de plaidoyer interne et externe » (Ramalingam et coll., 2009 : 64). Les recherches (souvent menées sous forme d'études de cas) et les évaluations de programmes monétaires se poursuivent et sont réunies par le Cash Learning Partnership (CaLP). Le site web du CaLP compte à présent 45 évaluations de programmes monétaires et 40 rapports de recherche.

Selon l'appréciation d'ALNAP, « la recherche et l'évaluation ont joué un rôle capital » dans l'acceptation de la programmation monétaire (Ramalingam et coll., 2009 : 63). Ses collègues de CaLP conviennent de l'importance des données. Ils considèrent

qu'une carence de données probantes dans certains domaines (ayant trait en particulier aux coûts et à l'efficacité des interventions monétaires par opposition à l'assistance en nature) constitue un obstacle à une plus grande acceptation de cette approche ; ils ont donc récemment créé un programme de recherche visant à combler ces déficits de données.¹⁹ En même temps, il est généralement admis que les données seules ne suffisent pas à dissiper les doutes et les réserves qu'éprouvent de nombreuses agences au sujet du recours à la programmation monétaire, notamment dans les situations d'urgence complexes. Dans un récent article, Degan Ali affirme : « bien qu'il ait existé des preuves de la viabilité et de l'efficacité des transferts monétaires » au sud-centre de la Somalie, « l'aversion au risque de la communauté humanitaire l'a rendue réticente à recourir à la programmation monétaire à grande échelle dès le début » ; en conséquence, « bien que cette approche se soit avérée efficace depuis longtemps dans la région, la décision (finale) de recourir à l'argent était davantage due aux personnalités favorables présentes et au manque d'alternatives qu'à une évaluation quelconque de l'efficacité et de l'adéquation des interventions monétaires pour répondre aux besoins de base » (Ali, 2012).

L'adoption de standards minimums

Buchanan-Smith examine l'importance des données de la recherche dans l'élaboration de la politique humanitaire dans un contexte assez différent. Dans son évaluation « Comment le projet Sphère a vu le jour », l'auteure s'intéresse à un changement précis : la décision d'adopter des standards volontaires minimums pour l'action humanitaire. Elle retrace le lien complexe entre l'Évaluation conjointe de l'aide d'urgence au Rwanda (Joint Evaluation of Emergency Assistance to Rwanda, JEEAR) et l'élaboration des standards Sphère. Elle en conclut que, si la JEEAR a eu « un très gros impact » (Buchanan-Smith, 2005 : 17) et a contribué à attirer l'attention sur le besoin d'établir des standards minimums, elle n'a en aucun cas été la seule cause des changements de politique, lesquels devaient « moins à la recherche, et plus aux inquiétudes croissantes dans le secteur humanitaire » (ibid. : 22). En effet, si une partie de la recherche issue de la JEEAR a eu de l'influence, bon nombre de ses conclusions les plus importantes « furent esquivées et constamment éludées » (ibid. : 24). Nous examinons ci-dessous une des conclusions qui n'a pas entraîné de changement.

Les exemples de la programmation monétaire et des standards minimums permettent de penser que les données peuvent contribuer – et contribuent – à l'élaboration de politiques dans le secteur humanitaire, mais que « le modèle qui considère l'élaboration de politiques comme un processus rationnel par lequel des données sont recueillies et des mesures appropriées recommandées est hautement invraisemblable » (Clarke et Ramalingam, 2008 : 32). Les évaluations, par exemple, sont « importantes, mais ne constituent qu'une seule des ressources et des influences en faveur du changement. [Elles figurent] en général au milieu du classement en termes de... leur intérêt pour le décideur » (Sandison, 2006 : 3).

L'élaboration des politiques ne repose pas exclusivement sur des données probantes, et celles-ci ne servent pas toujours à mettre sur pied des politiques. Nos deux derniers exemples illustrent l'échec du système humanitaire à effectuer des changements stratégiques ou politiques à partir de données : la prise en compte

“

L'élaboration des politiques ne repose pas exclusivement sur des données probantes, et celles-ci ne servent pas toujours à mettre sur pied des politiques

”

19 Voir <http://www.cashlearning.org/what-we-do/research-focus>

des facteurs culturels dans la programmation humanitaire, et l'intervention précoce dans les situations d'urgence liées à la sécheresse.

Prise en compte des facteurs culturels dans la programmation humanitaire

La non-prise en compte de facteurs culturels figure parmi les conclusions de la JEEAR et est une question qui semble avoir été « constamment éludée » selon Buchanan-Smith (2005 : 25).

Au Rwanda, l'équipe d'évaluation a constaté que « dans l'ensemble, les organismes de secours n'avaient qu'une compréhension très limitée de la structure de la société rwandaise et qu'il avait été fait peu cas du point de vue des bénéficiaires... un grand nombre d'employés de ces organismes n'avaient pas travaillé dans la région auparavant, connaissaient peu la société rwandaise et, de ce fait, ignoraient de nombreuses questions préoccupantes pour les Rwandais ordinaires parlant le kinyarwanda (JEEAR : 1996a : 176).

Le manque de connaissances contextuelles entraîna une série d'erreurs qui réduisirent l'efficacité, l'efficience et la pertinence de l'intervention : distribution de biens inappropriés, distribution de biens via des mécanismes communautaires qui excluaient les populations vulnérables et permettaient aux agents de l'État de se bâtir une assise de pouvoir qui contribuait à l'insécurité, et soutien à une politique de rapatriement précoce.

Reconnaissant que beaucoup de ces erreurs n'étaient pas commises par des ONG qui avaient l'expérience de l'action au Rwanda, l'évaluation concluait qu'il était « impératif que les ONG actives dans des situations d'urgence complexes : engagent du personnel qualifié qui a une expérience professionnelle dans de tels milieux et qui est conscient du besoin d'être sensible à la culture locale ; établissent des partenariats avec les organismes locaux [et] ; incluent au moins certains agents ou conseillers dotés d'une expérience considérable dans le pays » (JEEAR : 1996b : 61).

Au cours de la décennie suivante, ces conclusions seront reprises dans la recherche universitaire qui signalait l'importance de comprendre le contexte local dans lequel une intervention d'urgence avait lieu et de prendre en compte les perceptions des populations locales (Dijkzeul, 2010). Les évaluations sont revenues régulièrement sur ce thème (se reporter à Ali et coll., 2005 ; Oxfam, 2004 ; World Vision, 2011 ; Nicholson et Desta, 2010 ; Tache, 2010).

Pourtant, en 2006, plus de dix ans après l'intervention au Rwanda, le rapport de synthèse de la Tsunami Evaluation Coalition signalait toujours « la mise à l'écart... [des] organismes locaux ; ... le remplacement de personnel local compétent par des équipes internationales mal préparées ; la domination de l'anglais comme lingua franca ; ... l'imposition aux prétendus partenaires nationaux et locaux de conditions plus exigeantes que celles admises par les organisations internationales ; ... et une faible participation des bénéficiaires », autant de comportements répandus qui se soldaient par des « inégalités, des programmes insensibles aux questions de genre et aux conflits, des indignités, des affronts culturels et du gaspillage » (Telford et Cosgrave : Synthèse 2006 : 93-4).

Ainsi, la situation ne semble pas s'être beaucoup améliorée depuis la réponse au tsunami. Seize ans après la JEEAR, et malgré les données évaluatives et scientifiques qui soulignent le besoin d'un changement, les bénéficiaires continuent de se sentir inadéquatement consultés (ESH, 2012). Très peu d'agents nationaux sont promus à des postes opérationnels (Buchanan-Smith et Scriven, 2011). Les ONG locales et les organisations de la société civile sont souvent marginalisées dans les opérations de secours (ESH, 2012). Enfin, la rotation du personnel international demeure élevée, empêchant les décideurs de comprendre de manière approfondie le contexte, social, économique et politique dans lequel ils opèrent (Bhattacharjee et Lossio, 2011 ; Currion, 2010 ; Darcy, 2012).

Intervention précoce dans les situations d'urgence liées à la sécheresse

Ceci ne représente pas le seul échec à trouver des réponses politiques solides aux situations où les données semblent indiquer qu'un changement est requis. Dans notre dernier exemple, nous examinons la façon dont le système humanitaire international réagit à la sécheresse dans les zones pastorales.

L'intervention humanitaire dite traditionnelle dans les sécheresses s'est caractérisée par des distributions de denrées alimentaires à grande échelle, déclenchées en général par des niveaux intolérablement élevés de malnutrition. Toutefois, au cours des vingt dernières années, des voix se sont élevées pour passer à un modèle alternatif de « réponse initiale » selon lequel les agences réagissent à l'alerte précoce à la sécheresse en lançant une série d'interventions ciblant les moyens d'existence, apportant un appui à la santé des troupeaux et maximisant les revenus découlant des ventes de bétail.

Dans leur étude « Money to burn » (De l'argent à gaspiller, 2012), Catley et Cullis font observer que ces démarches avaient été utilisées au Sahel et dans la Corne d'Afrique dans les années 1980 et 1990. En 2001, Aklilu et Wakesa, analysant la réponse à la sécheresse de 1999-2001 au Kenya, tirent la conclusion suivante : « Le cadre politique de la réponse à la sécheresse doit être repensé... aller au-delà du secours alimentaire... soutenir et maintenir, non pas les populations elles-mêmes, mais leur capacité à échanger et à faire vivre leurs élevages » (2001 : 33). Quatre ans plus tard, Jaspars parvient à des conclusions similaires, après avoir étudié les travaux existants et procédé à des études de cas (Jaspars, 2006).

Au fil du reste de cette décennie, une série d'autres évaluations et documents de recherche ont conforté l'idée selon laquelle une réorientation politique était requise (voir Sadler et coll., 2009 ; VSF, 2009 ; Burns et coll., 2008 ; ODI, 2006). Les données permettaient de penser que les réponses précoces étaient plus efficaces, plus acceptables pour les populations locales et considérablement plus efficaces économiquement (Abebe et coll., 2008).

Ces études et les programmes sur lesquels elles reposaient suscitèrent une réponse limitée. Une étude antérieure d'ALNAP sur ce sujet relève que certains bailleurs ont mis en place un financement pluriannuel et des mécanismes de financement souple afin de permettre de lancer des interventions pertinentes sans avoir besoin de demander de nouveaux fonds (Hedlund et Knox Clarke, 2011). Néanmoins, ces initiatives sont « généralement modestes et ne correspondent pas aux besoins des populations affectées » (ibid : 6).

Au cours de la dernière décennie, les interventions sur les moyens de subsistance ont de manière générale été sous-financées (HPG, 2006) ; elles n'ont pas été traitées comme des priorités pour le Fonds central d'intervention d'urgence (CERF) (Pantuliano et Wakesa, 2008) ; et elles n'ont représenté en Éthiopie que 2,2 % du financement total en faveur des secours pour la sécheresse en 2011 (Catley, 2012). En 2012, 11 ans après l'appel d'Aklilu et Wakesa à repenser les réponses aux sécheresses, un rapport financé par DFID préconise que « l'intervention précoce et les mesures de renforcement de la résilience doivent être la réponse prioritaire majeure aux catastrophes (Venton et coll., 2012). Selon le même rapport, si la communauté internationale avait fait appel au déstockage comme option de défaut au cours de ces dix ans, quelque 11 milliards de dollars auraient été économisés. Surtout, « si une réponse précoce avait pu sauver ne serait-ce qu'un petit nombre de... vies [perdues à la suite de la sécheresse de 2010-2011], des milliers d'enfants, de femmes et d'hommes seraient encore en vie » (Save the Children et Oxfam, 2012 : 13).

S'attaquer aux freins à l'utilisation des données

Les obstacles à l'exploitation des données en vue d'élaborer une politique humanitaire sont bien entendu nombreux. La JEEAR et la TEC recensent les facteurs très réels qui ont un effet dissuasif sur la production de données indiquant qu'une agence, une intervention ou une démarche ont peut-être échoué. Or, sans cela, il devient difficile de constituer une série de données. Dans bien des cas, les données sont dispersées et ne sont pas disponibles en un format unique, comparable (JEEAR, 1996 ; Redmond, 2010). S'agissant des interventions précoces, Levine et ses collègues notent que le nombre de programmes, et donc les éléments de base, reste assez limité (Levine et coll., 2011).²⁰ La pénurie de données comparables ne suffit toutefois pas à expliquer le manque d'intérêt que les responsables politiques ont porté à la question de l'ignorance du contexte ou à celle de l'intervention tardive.

L'inaction relative de la communauté humanitaire en présence de données probantes s'explique aussi par le fait que beaucoup de problèmes sont tout simplement trop difficiles à résoudre. Autrement dit, il y a peu à gagner d'une connaissance accrue du contexte si d'autres facteurs rendent difficile, voire impossible, l'exploitation de ce savoir. Il s'agit notamment des facteurs suivants : « le manque de souplesse et la nature axée sur l'offre du système de secours international » (JEEAR, 1996a : 177) ; « les restrictions stipulées par les bailleurs sur la façon dont [les agences] utilisent les fonds » (Comité de pilotage TGLLP : 11), aspect de la dépendance au chemin emprunté discutée auparavant ; enfin, « l'urgence de dépenser l'argent de manière plus visible » (Telford et Cosgrave, 2006 : 93).

S'il y a sans doute plus qu'une parcelle de vérité dans le rôle de ces facteurs, cela n'explique là encore pas tout. Certains changements stratégiques visant à nouer des partenariats à long terme avec les acteurs de la société civile, à assurer que davantage de personnel recruté localement est affecté à des postes de prise de décision, ou à réduire la rotation du personnel dans les situations d'urgence, semblent viables et contribueraient à garantir que les programmes reposent sur des connaissances

20 Andy Catley propose une autre raison intéressante pour laquelle le corpus de données formel pour l'intervention précoce pourrait être modeste : l'approche est fondée sur une « logique causale » si forte que les praticiens n'ont pas ressenti la nécessité de tester la supposition de manière formelle – ce qui soulève la question intéressante du rôle des cadres logiques dans la fourniture de données.

contextuelles plus solides. Les bailleurs pourraient débloquer davantage de fonds plus tôt dans les catastrophes à survenue lente, et les agences pourraient être mieux préparées à intervenir. Comme le montre l'exemple de la programmation monétaire, les données probantes peuvent encourager des politiques qui remettent en question certains éléments existants du modèle humanitaire – dans le cas de l'argent liquide, l'idée que « l'argent n'était pas une option viable parce que l'on ne pouvait pas faire confiance à ses bénéficiaires pour le dépenser efficacement » (Ramalingam et coll., 2009 : 44). Alors pourquoi les données semblent-elles si souvent négligées ? Et que faut-il faire pour que ces données soient utilisées ?

Les difficultés liées à l'exploitation des données pour élaborer des politiques ne sont pas seulement d'ordre pratique : selon Sandison, il est particulièrement ardu de prendre des mesures, instrumentales ou générales, à partir d'évaluations qui « remettent en question des convictions et des comportements profondément enracinés dans la culture d'une organisation » (Sandison, 2006 : 111). L'auteur conclut que « la question de l'exploitation des évaluations relève autant de l'humain que du technique » (ibid : 132). Clarke et Ramalingam, dans leur étude du changement dans les organisations humanitaires, observent que « les personnes interrogées évoquaient des 'réactions viscérales' à des changements qui, en apparence, étaient plutôt de nature simplement technique » (2008 : 45).

Pour être efficace, le changement – y compris l'élaboration et l'adoption de nouvelles politiques – doit engager un processus qui réponde aux besoins rationnels, politiques et émotionnels des parties prenantes au sein de l'organisation. Reconnaissant cela, le programme Recherche et politique dans le développement (RAPID) de l'ODI a mis en place un cadre de travail pour examiner l'influence de la recherche sur les politiques (Young et Court, 2004). Ce cadre s'intéresse non seulement à la crédibilité et à la communication des informations émanant de l'évaluation, mais aussi aux liens entre les évaluateurs, les responsables politiques et les autres réseaux, au contexte politique ainsi qu'à l'influence de l'environnement extérieur. Ces facteurs semblent aussi importants pour déterminer la mesure dans laquelle les données probantes influencent la politique humanitaire.

Le cadre RAPID souligne l'importance de la communication des données aux décideurs. Dans le cas des transferts d'argent, l'étude d'ALNAP a constaté que « l'exploitation des résultats [des évaluations] de manière simple et forte... était capitale » (Ramalingam et coll., 2009 : 3). Comme on l'a vu, si des programmes monétaires existaient depuis un certain temps, c'est grâce au rapport *Buying power* ("Pouvoir d'achat") de la Croix-Rouge britannique en 2000 que les résultats de cet organisme furent pour la première fois aisément accessibles aux responsables politiques.

HPG et plus tard CaLP ont contribué à assurer que les données probantes et les apprentissages sont disponibles et recueillis. De même, la JEEAR « a clairement exposé et analysé ce que la plupart des agences humanitaires savaient déjà » (Buchanan-Smith, 2005 : 22) et a bénéficié d'un processus de suivi financé qui a permis à l'équipe d'évaluation de « vendre » le rapport et les messages clés qu'il contenait. En revanche, le peu d'attention accordée aux questions de contexte pourrait découler, en partie, d'un manque de synthèse claire des données : « si les perceptions de l'action humanitaire au niveau local sont de plus en plus prises en compte, elles ne l'ont pas été de manière assez systématique... ces études font rarement référence les unes aux autres » (Dijkzeul et Wakenge, 2010 : 1146). Vu l'importance de rendre

les données accessibles, les efforts déployés par des agences comme Oxfam, Care et NORAD pour synthétiser et rendre publiques les fruits de la recherche et des évaluations (Hallam, 2011) et par des groupes et réseaux tels qu'ALNAP peuvent tous contribuer à un système plus axé sur des données probantes (Dijkzeul et coll., 2012).

Il semblerait cependant que l'accès à l'information ne soit pas suffisant pour en garantir l'utilisation. Selon plusieurs études du milieu humanitaire, à tous les échelons les décideurs humanitaires ont tendance à être fortement influencés par les attitudes et les opinions de leurs pairs (Clarke et Ramalingam, 2008 ; ALNAP, 2003 ; Sandison, 2007 ; Darcy, 2009 ; Maxwell et coll., 2013). Cette constatation laisse penser que, dans le secteur humanitaire, les connaissances sont socialement construites et validées, et que pour que des données soient utilisées, il faut qu'elles soient avant tout incorporées au discours humanitaire.



Dans le secteur humanitaire, les connaissances sont socialement construites et validées.



Le rôle important que jouent les réseaux et les relations pour rendre les savoirs « acceptables » a été relevé ailleurs. Latour a démontré « comment la construction scientifique est une entreprise sociale dans laquelle convaincre les personnes à accepter certaines vérités dépend plus des relations sociales que du recours à des méthodes scientifiques » (cité dans Hilhorst, 2003). Jones et Mendizabal, eux, affirment que « les relations directes interpersonnelles entre le personnel et à la fois les chercheurs et les évaluateurs... jouent un rôle important » dans l'exploitation des données (Jones et Mendizabal, 2010).

Cette tendance, qui a des implications inquiétantes pour la capacité des savoirs « locaux » à influencer le cours de l'action humanitaire, a contribué à faire accepter l'argent liquide comme outil de programmation. Ramalingam et ses collaborateurs (2006 : 55) affirment en effet que « l'apparition d'un groupe dispersé ayant l'expérience du terrain qui a commencé à étudier... les possibilités de la programmation monétaire, et à combattre la méfiance des sceptiques », et qui a donné le jour à l'initiative d'apprentissage des interventions monétaires après le tsunami de 2004 (et, par la suite, à CaLP), a été importante pour faire accepter cette approche. Cet élan social, allié aux données issues des évaluations, demeure capital pour un développement pérenne des programmes de transferts monétaires.²¹

De même, la JEEAR a été une initiative interagences qui a mobilisé un large éventail d'acteurs provenant de l'ensemble du système humanitaire : la recherche était dès le départ sociale par sa nature. En revanche, Levine et ses collègues (2011) indiquent qu'il n'existe aucun forum de discussion de l'intervention précoce, et que la plupart des échanges sont bilatéraux et ont trait à des programmes précis : le réseau social autour de l'intervention précoce ne semble pas exister.

Le cadre RAPID souligne par ailleurs le rôle déterminant des politiques (organisationnelles) et de la pression extérieure dans le degré d'utilisation des données. Les responsables des politiques humanitaires sont sélectifs ; ils filtrent les données et « prennent en fin de compte les décisions concernant les recommandations des chercheurs en matière de changement de politique qu'ils sont disposés à accepter » (Buchanan-Smith, 2005). Dès lors, le système humanitaire... réagit plus au changement lorsqu'il est sous pression, lorsque les facteurs d'impulsion sont forts » (ibid : 98).

²¹ Selon certains membres du forum de discussion CaLP, qui ont eu la gentillesse de nous donner leur avis sur ce sujet.

Ces derniers étaient particulièrement robustes au lendemain de l'intervention au Rwanda, par exemple, les agences participant à des débats internes pour trouver des moyens d'améliorer leur performance et les financeurs exigeant des mesures. Il ne fait donc aucun doute que cela a permis d'assurer que les données probantes fournies par les équipes de la JEEAR ont au moins été entendues.

Certaines recommandations – notamment celles concernant les standards – ont ensuite été filtrées. Dans le cas des programmes monétaires, leur développement a eu lieu sur fond de vieilles inquiétudes quant aux effets de l'aide alimentaire et a été influencé de l'extérieur par divers facteurs : augmentation massive du financement après le tsunami, soutien des gouvernements de la zone de l'océan Indien en faveur des programmes monétaires, et hausse mondiale des cours des denrées alimentaires et du pétrole en 2008, qui a rendu plus chère la distribution de l'aide alimentaire.

Dans le cas de l'intervention précoce, en revanche, des facteurs tant organisationnels que politiques semblent œuvrer à l'encontre du changement de politique. Il est possible que de nombreuses agences évitent la programmation portant sur les moyens de subsistance parce que les compétences et les connaissances nécessaires leur font défaut (Aklilu et Wakesa, 2001), et parce que le modèle « sapeurs-pompiers » qui consiste à établir une présence dans une zone lorsqu'une catastrophe survient n'est pas efficace pour l'intervention précoce (HPG, 2006). En outre, certaines organisations reçoivent des fonds conséquents provenant de la monétisation de (vastes quantités) d'aide alimentaire requise pour faire face à des conditions critiques (Jaspars, 2006) : les sommes relativement limitées requises pour le soutien aux moyens de subsistance ne fourniraient pas les mêmes niveaux de recettes.

De plus, les bailleurs sont parfois réticents à intervenir sur la base de la seule prédiction, exigeant des données concrètes avant d'engager l'argent du contribuable (Save the Children et Oxfam, 2012 ; Levine et coll., 2011) et ils sont souvent incités à dépenser des sommes d'argent plus élevées que ce que demandent les ONG pour les interventions précoces. Ainsi, selon Levine et ses collègues, le représentant d'un bailleur aurait déclaré : « les ONG prennent de petites sommes d'argent... si nous donnons un gros chèque à l'ONU, nous pouvons le déduire de nos dépenses tout de suite (2011 : 7). D'importantes contraintes existent aussi au niveau politique : « les gouvernements nationaux considèrent souvent la déclaration d'un état d'urgence comme un signe de faiblesse » (Save the Children et Oxfam, 2012). Par conséquent, ils retardent l'alerte à l'urgence jusqu'à ce qu'il soit trop tard pour que les stratégies portant sur les moyens d'existence puissent être particulièrement efficaces (Hedlund et Knox Clarke, 2011).

La traduction des données probantes en politiques suit rarement un cheminement droit et clair. Dans certains cas, notamment lorsqu'elles ébranlent les idées reçues ou les démarches habituelles, les données risquent bel et bien de tomber complètement à l'eau. Dans bien des cas toutefois, les données probantes informent, voire orientent effectivement les politiques : c'est le cas en particulier lorsque les données deviennent généralement acceptées, soit grâce au soutien de coalitions ou d'autres structures, soit sous l'influence d'une crise qui rend les personnes plus réceptives à l'idée de changement.

EN BREF :

- Il existe une différence entre l'exploitation directe, « instrumentale » des évaluations et le rôle indirect à plus long terme de ces dernières pour influencer la politique humanitaire.
- Les évaluations sont souvent utilisées comme instruments (de remaniement immédiat des programmes ou projets évalués), bien que cette utilisation soit sélective.
- Les recommandations difficiles à mettre en œuvre ont moins de chances d'être suivies
- ALNAP a étudié les facteurs qui mènent à une utilisation instrumentale : le plus important d'entre eux semble être l'implication étroite des décideurs dans le processus d'évaluation ; de nombreux organismes humanitaires font désormais appel à des méthodes qui favorisent ce type d'implication.
- Le rôle des évaluations (et de la recherche évaluative) dans l'influence des politiques est envisagé en prenant quatre exemples : les programmes monétaires ; la conception de programmes pour leur adéquation au contexte ; l'évolution des standards humanitaires ; et l'intervention précoce dans les situations de sécheresse.
- Les évaluations ne représentent qu'une source de données probantes pour l'élaboration des politiques, et leur influence est souvent limitée par le fait qu'elles peuvent être difficiles à trouver (en partie du fait de la réticence de la part des agences à diffuser les données qui ont une incidence négative sur leurs programmes), et par la résistance naturelle au changement affichée par les individus et les organisations du secteur humanitaire.
- De plus, bon nombre d'évaluations indiquent des changements qui nécessitent une réorganisation fondamentale de l'architecture humanitaire et du « modèle commercial » de l'action humanitaire.
- Les évaluations ont plus de chances d'informer les politiques dans des situations où une pression pour le changement se fait déjà ressentir.
- Les forums qui réunissent divers acteurs et permettent de « socialiser » les connaissances acquises grâce aux évaluations sont eux aussi importants pour favoriser l'exploitation des évaluations.

5

Conclusions et recommandations : Comment peut-on améliorer la qualité et l'exploitation des données dans l'action humanitaire ?

5.1 Améliorer la qualité des données

Des données de qualité – et notamment des données sur la nature des besoins humanitaires et sur le succès ou l'échec des interventions – forment une composante essentielle de la programmation humanitaire. En conséquence, des processus de recueil et d'analyse de l'information sont en place à plusieurs stades du cycle de la programmation humanitaire standard.

Une évaluation de la qualité des données probantes générées par ces processus (au moyen des critères de données « valables » exposés à la section 2) laisse penser que beaucoup de progrès restent à faire. Le présent bilan semble aussi indiquer que des améliorations sont possibles : si les contextes humanitaires présentent de sérieuses entraves à la production de données probantes, beaucoup d'organisations du système humanitaire trouvent des moyens de les surmonter.

Sur la base de ce bilan, nous proposons que les efforts d'amélioration de la qualité des données dans le secteur humanitaire prennent en compte cinq principes clés.

MÉTHODOLOGIE SOLIDE : Une grande partie des données produites dans le secteur sont obtenues via des méthodes qualitatives. Toutefois, ces méthodes sont souvent mal comprises et mal appliquées. Les agences peuvent – et doivent – améliorer leur rigueur méthodologique à cet égard. En même temps, certaines agences expérimentent des méthodes quantitatives de collecte et d'analyse des données. Ces initiatives ont démontré à la fois les possibilités et les limites de ces démarches, et il convient de les encourager. Les tentatives en vue d'adopter des approches méthodologiques mixtes en matière de production de données doivent être encouragées.

Tous les rapports d'alerte précoce, d'appréciation, d'évaluation et de recherche produits par les acteurs humanitaires doivent être sans équivoque sur la solidité des données probantes utilisées, et contenir une description de la méthodologie employée pour collecter et analyser l'information, ainsi que les limites de cette méthodologie.

Les organisations devraient mettre en commun leurs expériences et leurs enseignements sur la façon dont elles utilisent les méthodologies qualitatives et quantitatives, de manière bilatérale et via des réseaux et des groupes interorganisations tels qu'ALNAP, ACAPS, et le Groupe de travail sur l'évaluation des besoins de l'IASC.

Ces réseaux devraient synthétiser les pratiques de références et favoriser la production de matériel de formation et d'autres outils de diffusion de ces pratiques.

Les organisations humanitaires devraient envisager de créer des organismes chargés du contrôle de la qualité de leurs rapports.

Les organisations humanitaires doivent déterminer si elles ont la capacité d'effectuer la collecte et l'analyse des données de manière pertinente par rapport à leur action et, si tel n'est pas le cas, elles devraient combler ces lacunes par le biais de partenariats stratégiques, d'embauches et de formation.

Les acteurs humanitaires devraient tisser des partenariats avec les universitaires, et s'inspirer de leurs enseignements, afin d'améliorer le niveau de qualité de la production de données.

EFFICIENCE : Les données probantes sont des apports à la programmation humanitaire, et en tant que tels le temps et l'argent investis pour leur production doivent être considérés dans le contexte des dépenses globales au titre des programmes actuels et futurs.

Les bailleurs et les organisations humanitaires doivent clairement prendre en compte les bénéfices potentiels – directs et à plus long terme – lorsqu'ils investissent dans la collecte de données, et être disposés à investir des sommes proportionnelles aux économies ou aux améliorations qui pourraient être réalisées.

COLLABORATION : L'un des principaux obstacles à la production efficace de données de haute qualité (et à leur exploitation) dans le système humanitaire est le manque de collaboration au sein des organisations et entre celle-ci. En effet, les données sont souvent recueillies en silos fonctionnels ou organisationnels. Ce système empêche toute vérification croisée et tout contrôle qualité rigoureux, encourage la duplication des activités et empêche effectivement la plupart des acteurs humanitaires d'accéder à de vastes quantités de données. Les initiatives qui visent à produire des données probantes humanitaires doivent reconnaître leur responsabilité de contribuer à un corpus de données sectoriel.

Les bailleurs doivent faire appel aux réseaux et aux forums interagences afin d'identifier les lacunes de données qui existent au sein du système humanitaire, et élaborer des stratégies communes pour les combler.

Tous les acteurs humanitaires, et notamment les bailleurs, devraient continuer à encourager les processus d'identification de cadres standardisés globaux pouvant être utilisés dans le cadre de l'alerte précoce, de l'appréciation et de l'évaluation conjointes ou coordonnées.

Les EHP, clusters et mécanismes de coordination inter-cluster doivent clarifier les besoins de données à l'échelon national et encourager les appréciations et les évaluations conjointes ou coordonnées.

Toutes les activités de recueil de données doivent inclure des budgets et des plans afin d'assurer la diffusion des données, et maximiser l'appui des organes opérant à l'échelle du système (ACAPS, ALNAP et CaLP, entre autres) pour favoriser la diffusion.

CONTINUITÉ : Les cycles de financement courts et la rotation rapide du personnel tendent à œuvrer à l'encontre de la collecte d'ensemble de données comparables dans le temps. En conséquence, les données sont souvent produites de manière épisodique, à l'occasion d'évaluations ponctuelles. Cela réduit considérablement notre capacité à cerner les effets de l'intervention humanitaire dans le temps, et, partant, nous empêche de dire, en toute confiance, quand et comment nous devrions intervenir à l'avenir. Les acteurs humanitaires doivent envisager la production de données comme un processus et non comme un événement, et viser à constituer un corpus de données à long terme.

Les humanitaires devraient se garder de remanier constamment leurs systèmes d'évaluation et de suivi. Les systèmes d'évaluation et de suivi doivent être conçus de sorte à se compléter les uns et les autres. Si les systèmes sont remaniés, il y a lieu de veiller à continuer de recueillir les données concernant les indicateurs clés.

Les humanitaires doivent soutenir l'action menée à l'échelon national afin de créer des dispositifs de suivi et d'autres systèmes capables de fournir des données de référence, et des séries chronologiques plus étendues de données pour les indicateurs clés, la réduction des risques de catastrophes et les plans d'urgence.

INCLUSION ET APPROPRIATION : Les personnes qui sont les plus étroitement concernées par l'action humanitaire, celles qui sont personnellement touchées par une crise, ont tendance à être écartées des méthodes actuelles de production de données de deux manières. Premièrement, elles n'ont pas la possibilité de poser les questions : les données sont généralement recueillies afin de répondre aux besoins des organisations internationales plutôt qu'à ceux de la population affectée. Les évaluations, par exemple, ont tendance à collecter des données qui permettent de savoir quel type d'aide est requis, plutôt que de déterminer si une aide est nécessaire en premier lieu. Deuxièmement, leurs réponses sont souvent considérées comme sans importance : la présente étude a indiqué que même lorsque des données sont réunies pour déterminer si une situation constitue une crise, ou si une intervention précise s'est avérée efficace, les connaissances et les opinions des personnes les plus directement affectées ne recevaient qu'une attention limitée. À l'avenir, les organisations internationales devront expliquer clairement pourquoi elles recueillent des données et à qui celles-ci sont destinées. Elles devront examiner dans quelle mesure elles peuvent réunir des données utiles pour les organisations de la société civile, et redoubler d'efforts pour inclure les voix des populations affectées dans leur collecte de données.

Les organisations internationales doivent recenser les besoins en informations et en données ressentis par les organisations de la société civile dans les régions touchées par des catastrophes et fournir des informations afin de répondre à ces besoins. Le cas échéant, elles doivent également proposer des formations et des ressources afin de permettre aux organisations nationales et locales de recueillir l'information.

Les rapports d'alerte précoce, d'appréciation, d'évaluation et de recherche produits par les acteurs humanitaires doivent exposer clairement les savoirs et les propositions locaux. Lorsque les recommandations ne correspondent pas à ces informations ou propositions locales, les auteurs doivent donner la raison de cette différence.

Les informations doivent, dans tous les cas, être recueillies sur la base d'un consentement informé. En outre, les organisations qui recueillent les données doivent trouver des moyens d'informer la population sur la manière dont elles ont été utilisées et sur les décisions éventuelles qui ont été prises à partir de celles-ci.

5.2 Améliorer l'utilisation des données probantes dans la prise de décision humanitaire

Le bilan a également étudié dans quelle mesure les décideurs au sein du système humanitaire tiennent compte des données lorsqu'ils prennent leurs décisions. En général, nous avons conclu que des améliorations considérables étaient nécessaires à cet égard. En effet, dans bien des cas, les processus de prise de décision n'ont pas pris en compte des données probantes importantes fournies par les systèmes d'alerte précoce, les appréciations ou évaluations des besoins, ce qui a eu des répercussions néfastes sur les vies et les moyens de subsistance de la population touchée par une catastrophe. C'est le cas notamment lorsque les données remettent en question des interprétations préconçues sur une situation ou font ressortir un besoin de changement radical.

Pour autant, il y a lieu de ne pas représenter de manière trop simpliste la façon dont les données peuvent ou doivent être exploitées dans la prise de décision. En effet, la plupart des décisions prises dans le secteur humanitaire requièrent la prise en compte d'une multitude d'ensembles de données qui sont souvent rassemblées à l'aide de méthodologies différentes et indiquent des plans d'action différents. Dans de telles circonstances, il n'est ni réaliste, ni souhaitable d'espérer que les données déterminent à elles seules la décision. Il y a toujours une part nécessaire de jugement dans la prise de décision et nous devons nous efforcer d'assurer que ces jugements sont, dans tous les cas, informés par les meilleures données disponibles.

Améliorer l'exploitation des données – collecte, analyse et diffusion

Compte tenu de ce bilan, nous estimons que les données ont plus de chances d'influencer la prise de décision si les personnes responsables de la production de ces données sont guidées par les principes suivants :

ACCESSIBILITÉ : Les décideurs ont rarement le temps (voire parfois l'envie) d'explorer des bases de données, des rapports ou des tableurs pour obtenir les informations dont ils ont besoin pour prendre une décision solide. Aussi les données doivent-elles être présentées de manière à être aisément accessibles, c'est-à-dire sous la forme de rapports courts rédigés en termes simples, de graphismes d'information et de portails interrogeables.

Les réseaux d'apprentissage humanitaire (tels qu'ALNAP), les initiatives interagences (tels que les clusters internationaux) et les organisations spécialisées doivent faire tout leur possible pour assurer que les informations et les données sont aisément accessibles par les moyens suivants :

- synthèse des pratiques de référence et des enseignements tirés des évaluations et de la recherche ;

- création et maintenance de plateformes aisément accessibles et interrogeables pour les données documentaires et autres, les entités qui créent des bases de données veillant à ce qu'elles collaborent avec les initiatives déjà en place afin de prévenir toute duplication et confusion.

RAPIDITÉ : Toute information qui arrive après que les décisions ont été prises est souvent ignorée, et les décisions les plus importantes – celles qui fixent le champ d'action global d'un programme – sont souvent prises très tôt. Les groupes ou individus qui produisent des données doivent viser à comprendre le processus de prise de décision et à rendre les données pertinentes disponibles aux points critiques du processus.

Les organisations humanitaires responsables de l'alerte précoce et de l'évaluation doivent déterminer dans quelle mesure ces instruments sont produits au bon moment pour influencer les décisions.

Le cas échéant, elles doivent tenter de modifier leurs procédures pour assurer que les informations sont disponibles en temps utile.

LARGE CIRCULATION VIA DIVERS MÉDIAS : Dans le secteur humanitaire, les connaissances semblent en grande partie socialement construites. Les décideurs sont influencés par les connaissances et les convictions de leurs pairs. Dès lors, des informations largement diffusées et largement connues et débattues ont bien plus de chances d'être exploitées comme données en vue d'étayer ou de rejeter certaines mesures.

Les organisations qui produisent des données doivent assurer leur diffusion active via un éventail de méthodes : webinaires, communautés de praticiens, réunions, vidéos, stages de formation, etc.

Grâce à ces divers médias, elles doivent viser à impliquer divers groupes, notamment les responsables politiques, les praticiens et les représentants du leadership organisationnel.

Les bailleurs qui soutiennent la production de données doivent veiller à ce que leurs budgets prévoient la diffusion des résultats.

Améliorer l'exploitation des données – prise de décision

S'il incombe aux humanitaires d'assurer que leurs données sont accessibles, disponibles en temps utile et largement diffusées, ces activités ne suffisent pas à elles seules à accroître la prise en compte des données probantes par les décideurs. Les organisations affiliées à ALNAP doivent également prendre des mesures afin d'améliorer leurs méthodes de prise de décision.

METTRE EN PLACE DES PROCÉDURES CLAIRES DE PRISE DE DÉCISION :

Les données ont plus de chances d'être exploitées si les organisations instaurent des procédures de prise de décision comprenant des mesures claires qui nécessitent la prise en compte des données.

Les organisations humanitaires doivent décrire clairement le processus selon lequel leurs programmes sont conçus (et remaniés), en indiquant les étapes

clés auxquelles les informations sont requises, et, le cas échéant, la nature des informations requises afin de prendre les décisions.

Les bailleurs doivent exiger des organisations humanitaires qui leur soumettent des propositions de financement qu'elles leur fournissent les données les plus probantes disponibles (compte tenu des implications en matière de temps et de ressources) afin de démontrer que :

- une assistance extérieure est nécessaire ;
- l'intervention proposée est la meilleure option disponible ; et
- l'organisation proposante a la capacité de mettre en œuvre l'intervention efficacement.

Il conviendrait que les agences humanitaires envisagent l'introduction de tableaux matriciels de suivi de l'intervention/décision (lorsqu'elles n'y ont pas déjà recours) pour permettre aux parties prenantes de voir dans quelle mesure l'organisme a donné suite aux recommandations issues de l'évaluation.

Dans le domaine de l'alerte précoce ou de l'intervention précoce, le système humanitaire doit créer une plate-forme commune visant à étudier l'ensemble du processus, de l'alerte précoce à l'action initiale, en mettant l'accent sur le rôle des systèmes d'alerte précoce dans l'information des décisions, la situation de l'autorité de prise de décision dans les agences d'intervention, et des apprentissages communs portant sur les activités d'intervention précoce.

VEILLER À CE QU'IL Y AIT DES MESURES INCITATIVES POUR L'UTILISATION DES DONNÉES : l'utilisation des données dans la prise de décision a souvent un coût en temps (lié à la recherche et au traitement des données) et peut aboutir à des décisions qui menacent les méthodes de travail préétablies, ou qui sont pour une autre raison « difficiles à mettre en œuvre ». Il pourrait être plus facile pour les décideurs de ne pas rechercher ou utiliser des données. Pour remédier à cela, les organisations devraient envisager des mesures incitatives pour l'utilisation des données.

Tout standard futur et toute procédure de certification assimilée utilisés dans le secteur humanitaire devraient exiger des agences qu'elles répondent à des normes agréées en matière de qualité de la collecte des données.

Les organisations humanitaires doivent clarifier les rôles et les attentes des décideurs en matière de prise de décision. Il convient de déterminer clairement à qui incombe la prise de telle ou telle décision : les responsables doivent être conscients des attentes de l'organisation en ce qui concerne l'exploitation des données.

Encadré 6 Que fera ALNAP ?

Nous espérons que les organisations membres d'ALNAP réfléchiront et mettront en œuvre ces recommandations. Entretemps, au nom du Réseau, le secrétariat d'ALNAP :

- œuvrera de concert avec la communauté de praticiens de l'évaluation en vue d'élaborer et de partager des orientations pratiques sur les manières d'améliorer la qualité des données issues des évaluations humanitaires ;
- développera du matériel de formation pour aider les membres d'ALNAP à améliorer la qualité probante des évaluations, et ce à partir du guide Evaluating Humanitarian Action (EHA) (Buchanan-Smith et Cosgrave, (2013) ;
- continuera à actualiser et à améliorer son Centre de ressources, en assurant qu'il abrite le plus grand nombre possible d'évaluations, documents d'enseignements et rapports sur l'action humanitaire et qu'il est accessible et facilite la recherche ;
- continuera à expérimenter avec une palette de méthodologies de recherche (analyses bibliographiques structurées, méthodes d'analyse d'études de cas par appariement, méthodes d'enquête quantitative, etc.) pour assurer que les rapports d'ALNAP se fondent sur les données probantes les plus solides possible : publication et diffusion de nos expériences à l'aide de ces méthodologies ;
- œuvrera avec les membres d'ALNAP (via nos travaux sur le leadership opérationnel) pour mieux comprendre les processus de prise de décision et d'évitement du risque ou de l'aversion de l'incertitude, et publiera des recommandations concrètes pour améliorer la prise de décision à partir d'une recherche solide.

Annexe 1 : Méthodologie

La version initiale du présent rapport, « Evidence and Knowledge in Humanitarian Action » (Données et savoirs dans l'action humanitaire) reposait sur une analyse bibliographique, laquelle était augmentée d'un nombre réduit d'entretiens. Elle a été diffusée en tant que document de référence à l'occasion de la 28^e réunion annuelle d'ALNAP et présentée à la réunion le 5 mars 2013.

Le rapport actuel s'inspire de ce document de référence, qui a été remanié afin d'incorporer les éléments suivants :

- les commentaires émis par les invités et les participants à la réunion à propos du document de référence lui-même ;
- les recommandations formulées par les participants à la réunion lors de la dernière session de la réunion le 6 mars 2013 ; et
- les informations découlant des 26 présentations faites lors de la réunion.

Les présentations qui ont été incorporées au présent rapport ont été retenues sur la base d'un appel ouvert aux praticiens et aux universitaires spécialisés dans l'humanitaire via la liste des membres d'ALNAP. Les intervenants devaient discuter les « nouveaux apprentissages et pratiques de référence émergentes dans la compréhension des données probantes et dans le recueil, l'analyse et l'exploitation des données dans l'action humanitaire » (ALNAP, 2013 : 1). Au total, 26 présentations ont été retenues sur 110 propositions. Toutes les présentations étaient en anglais.

La présente étude comprend également les résultats d'une analyse bibliographique supplémentaire informée par les thèmes qui ont été soulevés à la réunion annuelle. Elle a été soumise à un examen collégial par un groupe d'experts.

Les principales limites dans cette démarche ont trait au biais de sélection. Les analyses bibliographiques n'étaient pas fondées sur une recherche structurée des bases de données universitaires et n'ont pas fait appel à des critères formels d'inclusion ou d'exclusion (Hagen-Zanker et Mallett, 2013).

De même, s'il est moins susceptible d'être influencé par un auteur unique, le choix des présentations faites lors de la réunion annuelle et référencées ici est biaisé par le canal de sélection (le réseau ALNAP) et par le fait que toutes les présentations ont été faites en anglais.

Annexe 2 : Liste des personnes interrogées et des contributeurs

Les personnes suivantes ont bien voulu accepter de discuter des questions abordées dans le présent rapport en personne ou par téléphone, ou de répondre aux questions par email :

Miriam Ait Aissa, ACF

Ben Allen, ACF

Haley Bowcock, CaLP

Jeff Crisp, HCR

Andrew Catley, université de Tufts

Mary Dixon-Woods, université de Leicester

Merry Fitzpatrick, World Concern

Richard Garfield, CDC

François Grünewald, Groupe URD

Joyce Luma, PAM

John Mitchell, ALNAP

Lili Mohiddin, CaLP

Lars Peter Nissen, ACAPS

Jock Paul, UN-OCHA

Chloe Puett, université de Tufts

Anthony Redmond, université de Manchester

Cécile Salpeteur, ACF

Louise Shaxson, ODI

Sharon Truelove, British Red Cross

Vivien Margaret Walden, Oxfam

Bibliographie

- Abebe, D. et al. (2008) 'Impact of a commercial destocking relief intervention in Moyale district, southern Ethiopia', in *Disasters* 32(2): 167–189. www.alnap.org/resource/8200.aspx
- ACAPS (2010) 'Rapid Initial Needs Assessment for Haiti (RINAH)'. Port-au-Prince: ACAPS. www.alnap.org/resource/9811.aspx
- ACAPS (2011) 'Joint Rapid Assessment of the Northern Governorates of Yemen'. Sanaa: ACAPS. www.alnap.org/resource/9812.aspx
- ACAPS (2012) 'Technical Brief: Estimation of affected population figures'. Geneva: ACAPS. www.alnap.org/resource/9813.aspx
- ACAPS (2012) 'Coordinated Assessments in emergencies. What we know now : Key lessons from field experience'. Geneva: ACAPS. www.alnap.org/resource/7912.aspx
- ACF (2011) 'Learning review'. London: ACF. www.alnap.org/resource/7001.aspx
- Aklilu, Y. and Wekesa, M. (2001) 'Livestock and Livelihoods in Emergencies: Lessons Learnt from the 1999–2001 Emergency Response in the Pastoral Sector in Kenya'. Nairobi: OUA IBAR. www.alnap.org/resource/9815.aspx
- Alexander, J. et al. (2013) '2013 Humanitarian accountability report'. Geneva: HAP International. www.alnap.org/resource/8758.aspx
- Ali et al. (2005) 'Cash Relief in a Contested Area: Lessons from Somalia: HPN Network paper no. 50'. London: HPN/ODI. www.alnap.org/resource/7585.aspx
- Ali, D. (2012) 'A Deadly Delay: Risk Aversion and Cash in the 2011 Somalia Famine,' in Bailey, S. and Ridsdel, B. (eds), *New learning in cash transfer programming [special feature] Humanitarian Exchange* 54. London: HPN/ODI. 10–15. www.alnap.org/resource/9817.aspx
- Allen, B. (2012) 'Collecting and Reporting best practices from the field'. *Msg 1, thread 2 under Capacity Area 2: Evaluation policy and purpose, in ALNAP Evaluation capacities – Community of Practice* (internal forum) London: ALNAP. (Accessed: 9 September 2012) www.alnap.org/resource/9818.aspx
- ALNAP (2003) 'Learning by Field Level Workers,' in *ALNAP review of Humanitarian Action in 2003*. London: ALNAP. www.alnap.org/resource/5208.aspx

- ALNAP (2010) 'The State of the Humanitarian System – Assessing Performance and Progress: a Pilot Study'. London: ALNAP/ODI. www.alnap.org/resource/8746.aspx
- ALNAP (2012) 'The State of the Humanitarian System 2012 edition'. London: ALNAP/ODI. www.alnap.org/sohsreport
- ALNAP (2013) 'Evidence and Knowledge in Humanitarian Action – Call for Presentations'. London: ALNAP/ODI. www.alnap.org/resource/7162.aspx
- Anderson et al. (2012) 'Time to listen: Hearing people on the receiving end of international aid'. Cambridge, Mass: CDA Collaborative Learning Projects. www.alnap.org/resource/8530.aspx
- Argren, R. (2013) 'What goes up must come down: the challenges of getting evidence back to the ground'. Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March. www.alnap.org/resource/8015.aspx
- Ashdown, P. (2011) 'Humanitarian Emergency Response Review'. London: HERR. www.alnap.org/resource/6355.aspx
- Assessment Working Group for Northern Syria (2013) 'Joint Rapid Assessment of Northern Syria – Aleppo City Assessment'. Syria: Assessment Working Group for Northern Syria. www.alnap.org/resource/8755.aspx
- Avila et al. (2010) 'Global mapping of technology for transparency and accountability'. London: Open Society Foundation. www.alnap.org/resource/6421.aspx
- Banatvala, N. and Zwi, A. (2000) 'Public health and humanitarian interventions: developing the evidence base', in *The British Medical Journal* 321:101–5. www.alnap.org/resource/9820.aspx
- Beck, T. (2006) 'Evaluating humanitarian action using the OECD-DAC criteria – An ALNAP guide for humanitarian agencies'. London: ALNAP/ODI. www.alnap.org/resource/5253.aspx
- Bekele, G. and Abera, T. (2008) 'Livelihoods-based Drought Response in Ethiopia: Impact Assessment of Livestock Feed Supplementation', Pastoralist Livelihoods Initiative. www.alnap.org/resource/9936.aspx
- Bekele, G. (2010) 'Review of Save the Children US Livestock Marketing Initiative in Southern Ethiopia'. Westport: Save the Children USA. www.alnap.org/resource/9937.aspx
- Beynon et al. (2012) 'What difference does a policy brief make?'. Brighton: IDS and 3iE. www.alnap.org/resource/9938.aspx
- Bhattacharjee, A. and Lossio, R. (2011) 'Evaluation of OCHA Response to the Haiti Earthquake Final Report'. New York: UN OCHA. www.alnap.org/resource/6002.aspx

- Bickman, L. and Reich, S. (2009) 'Randomised Controlled trials: A gold standard with feet of clay?', in Donaldson et al. (2009) *What counts as credible evidence in applied research and evaluation practice*. Thousand Oaks: Sage.
www.alnap.org/resource/10028.aspx
- Bond. (n.d.). 'The NGO Evidence Principles'. London: Bond.
www.alnap.org/resource/9941.aspx
- Bonino, F. with Jean, I. and Knox Clarke, P. (forthcoming) Effective feedback mechanisms in humanitarian contexts - Guidance for practitioners and policy-makers based on findings from ALNAP-CDA case studies in Sudan, Pakistan and Haiti. ALNAP Study. London: ALNAP/ODI. www.alnap.org/resource/9944.aspx
- Borton et al. (1996a) '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
- Borton, J. et al. (1996b) 'The International Response to Conflict and Genocide: Lessons from the Rwanda Experience: Synthesis Report'. London: ODI.
www.alnap.org/resource/2517.aspx
- Bradt, D. A. (2009) 'Evidence-based decision-making in humanitarian assistance'. *HPN Network Paper 67*. London: HPN/ODI. www.alnap.org/resource/7503.aspx
- Broadbent, E. (2012) 'Politics of research-based policy in African policy debates: Synthesis of case study findings'. London: ebpdn.
www.alnap.org/resource/9942.aspx
- Buchanan-Smith, M. and Davies, S. (1995) 'Famine Early Warning Systems and Response: the Missing Link'. London: IT Publications.
www.alnap.org/resource/9943.aspx
- Buchanan-Smith, M. (2000) 'Role of Early Warning Systems in Decision-Making Processes'. London: HPN/ODI. www.alnap.org/resource/9946.aspx
- Buchanan-Smith, M. and Cosgrave, J. (2013) 'Evaluation of humanitarian action guide: Pilot guide'. London: ALNAP/ODI.
www.alnap.org/www.alnap.org/eha
- Buchanan-Smith, M. and Beck, T. (2008) 'Joint evaluations coming of age? The quality and future scope of joint evaluations', in *ALNAP Review of humanitarian action: Chapter three*. London: ALNAP/ODI. www.alnap.org/resource/5232.aspx
- Buchanan-Smith, M. and Scriven, K. (2011) 'Leadership in Action: Leading Effectively in Humanitarian Operations'. London: ALNAP/ODI.
www.alnap.org/resource/6118.aspx
- Buchanan-Smith, M. (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 et al. (eds) *Bridging Research and Policy in Development: Evidence and the Change Process*. London: ODI. www.alnap.org/resource/9947.aspx

- Burns et al. (2008) 'Impact Assessment of the Pastoralist Survival and Recovery Project, Dakoro'. Niger: Lutheran World Federation and Feinstein International Center. www.alnap.org/resource/5722.aspx
- CaLP (2013) 'Protecting Beneficiary Privacy: Principles and Operational Standards for the Secure Use of Personal Data in Cash and E-Transfer Programmes' Oxford: CaLP. www.alnap.org/resource/9750.aspx
- Carr, H. (2013) 'Tracking Beneficiaries and Services Delivered'. Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March. Washington: Catholic Relief Services. www.alnap.org/resource/8023.aspx
- Catley, A. (2013) 'Developing international standards and guidelines with limited hard evidence'. Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March. Washington: Tufts University. www.alnap.org/resource/8017.aspx
- Catley et al. (n.d.) 'Participatory impact assessment: A guide for Practitioners'. Boston: Feinstein International Centre and Tufts University. www.alnap.org/resource/8094.aspx
- Catley, A. and Cullis, A. (2012) 'Money to burn? Comparing the costs and benefits of drought responses in pastoralist areas of Ethiopia', in *The Journal of Humanitarian Assistance*. Ethiopia: FAO. www.alnap.org/resource/9948.aspx
- Clarke, P. and Ramalingam, B. (2008) 'Organisational change in the humanitarian sector' in *ALNAP 7th Review of Humanitarian Action*. London: ALNAP/ODI. www.alnap.org/resource/5235
- Cooley, A. and Ron, J. (2002) 'The NGO scramble: organizational insecurity and the political economy of transnational action', in *International Security* 27(1): 5–39. www.alnap.org/resource/9949.aspx
- Cosgrave, J. (2007) 'Tsunami Evaluation Coalition, Synthesis Report: Expanded Summary, Joint evaluation of the international response to the Indian Ocean tsunami'. London: ALNAP and TEC/ODI. www.alnap.org/resource/5536.aspx
- Coyle, D. and Meier, P. (2009) 'New Technologies in Emergencies and Conflicts: The Role of Information and Social Networks'. Washington: United Nations Foundation. www.alnap.org/resource/6572.aspx
- Currian, P. (2010) 'Coordination at the Crossroads: NGO coordination in Southern Sudan 2007–2011'. London: ALNAP and TEC/ODI. www.alnap.org/resource/5781.aspx
- Damerell, J. (2013) 'The role of standards for evidence based humanitarian response', Presentation at *Evidence in Humanitarian Action, ALNAP 28th Annual Meeting*. Washington: ALNAP. www.alnap.org/resource/8016.aspx

- DARA (2011) 'The Humanitarian Response Index 2011'. Madrid: DARA.
www.alnap.org/resource/6243.aspx
- Darcy, J. and Hofmann, C. (2003) 'According to need? Needs assessment and decision making in the humanitarian sector'. London: HPG/ODI.
www.alnap.org/resource/9384.aspx
- Darcy, J. (2009) 'Humanitarian diagnostics: the use of information and analysis in crisis response decisions'. London: HPG/ODI. www.alnap.org/resource/9485.aspx
- Darcy, J. and Garfield, R. (2011) 'Review of the Rapid Initial Needs Assessment for Haiti', paper prepared for ACAPS. Geneva: Assessment Capacities Project.
www.alnap.org/resource/9951.aspx
- Darcy et al. (2013) 'The Use of Evidence in Humanitarian Decision Making ACAPS Operational Learning Paper'. Boston: Feinstein International Center and Tufts University. www.alnap.org/resource/8003.aspx
- Development Initiatives (2012) 'Global Humanitarian Assistance: Tracking spending on cash transfer programming in a humanitarian context'. Wells: Development Initiatives. www.alnap.org/resource/9952.aspx
- de Ville de Goyet, C. and Moriniere, L. C. (2006) 'The Role of Needs of Assessment in the Tsunami Response'. London: ALNAP and TEC/ODI.
www.alnap.org/resource/3531.aspx
- DFID (2012) 'Promoting innovation and evidence-based approaches to building resilience and responding to humanitarian crises: a DFID strategy paper'. London: DFID. www.alnap.org/resource/9823.aspx
- DFID (2013) 'How to note – Assessing the Strength of Evidence'. London: DFID.
www.alnap.org/resource/8402.aspx
- DG ECHO (2009) 'The Use of Cash and Vouchers in Humanitarian Crises: DG ECHO funding guidelines'. Brussels: DG ECHO. www.alnap.org/resource/9824.aspx
- Dijkzeul, D. and Wakenge, C. (2010) 'Doing good, but looking bad? Local perceptions of two humanitarian organisations in eastern Democratic Republic of the Congo', in *Disasters* 34(4): 1139–70. www.alnap.org/resource/9953.aspx
- Dijkzeul et al. (2013) 'Introduction to Evidence based action in humanitarian crises'. [Manuscript submitted for publication], in *Disasters*, 37: S1–19.
www.alnap.org/resource/9954.aspx
- ECOSOC (2007) 'UNICEF evaluation policy'. New York: UNICEF.
www.alnap.org/resource/10037.aspx
- Epstein et al. (2004) 'Communicating evidence for participatory decision making' in *The Journal of the American Medical Association*, 291(19): 2359–66.
www.alnap.org/resource/9955.aspx

- Evans et al. (2011) 'Testing treatments: better research for better health care'. London: Pinter and Martin. www.alnap.org/resource/9956.aspx
- FAO and WFP (2013) 'Joint Evaluation of Food Security Cluster Coordination'. Rome: FAO/WFP. www.alnap.org/resource/9957.aspx
- Fearon et al. (2008) 'Community Driven Reconstruction in Lofa County: Impact assessment'. New York: International Rescue Committee. www.alnap.org/resource/8192.aspx
- Featherstone, A. (2011) 'Strength in numbers: a global mapping review of NGO engagement in coordinated assessments'. ECB Project. www.alnap.org/resource/9958.aspx
- Featherstone, A. (2013) 'Improving Impact: Do Accountability Mechanisms Deliver Results?'. London: Christian Aid, Save the Children for the Humanitarian Accountability Partnership. www.alnap.org/resource/8388.aspx
- Feinstein International Center (2007) 'Impact Assessments of Livelihoods-based Drought Interventions' in Moyale and Dire Woredas, Ethiopia: A Pastoralist Livelihoods Initiative report'. Boston: Feinstein International Center, Tufts University www.alnap.org/resource/3490.aspx
- Fenton et al. (eds) (2011) Humanitarian Accountability [special feature] *Humanitarian Exchange* 52. London: HPN/ODI. www.alnap.org/resource/6256.aspx
- Fetouh et al. (2013) 'Data quality in remote monitoring – a comparative analysis'. Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March. www.alnap.org/resource/8027.aspx
- Few, R. et al. (2013) 'Contribution to Change: An approach to evaluating the role of intervention in disaster recovery'. Oxford: Oxfam. www.alnap.org/resource/10305.aspx
- Foresti, M. (2007) 'A Comparative Study of Evaluation Policies and Practices in Development Agencies'. London: ODI. www.alnap.org/resource/7773.aspx
- Garfield, R. (2013) 'Evidence and needs assessments', Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March. www.alnap.org/resource/9959.aspx
- Garfield et al. (2011) 'Common Needs Assessments and Humanitarian Action'. *Humanitarian Network Paper* Number 69. London: HPN/ODI. www.alnap.org/resource/6262.aspx
- Gerring, J. (2011) *Social Science Methodology: A Unified Framework*. 2nd ed. Cambridge: Cambridge University Press. www.alnap.org/resource/9015.aspx
- Gilman, D and Noyes, A. (2013) 'Humanitarianism in the Network Age'. New York: OCHA. www.alnap.org/resource/9157.aspx

- Goldacre, B. (2012) *Bad pharma: how drug companies mislead doctors and harm patients*. London: Fourth Estate. www.alnap.org/resource/9960.aspx
- Gottwald, M. (2010) 'Competing in the humanitarian marketplace: UNHCR's organizational culture and decision-making process'. Geneva: UNHCR Policy Development and Evaluation Service. www.alnap.org/resource/9961.aspx
- Grünewald, F. (2012) 'Sustaining learning from evaluation: stakeholder engagement and 'customer care'?'. Msg 1, thread 2 under Capacity Area 3: Evaluation processes and systems. Message posted in *ALNAP Evaluation capacities – Community of Practice* (internal forum). London: ALNAP. www.alnap.org/resource/9818.aspx
- Grünewald et al. (2010) 'Inter-agency real-time evaluation in Haiti: 3 months after the earthquake'. Plaisans: Groupe URD. www.alnap.org/resource/9962.aspx
- Guenther et al. (2010) 'The politics of evaluation: evidence-based policy or policy-based evidence?' Paper presented to *the NARU Public Seminar Series*, 30 November 2010, Darwin. www.alnap.org/resource/9963.aspx
- Guerrero, S. (2012) 'Our vision of linking evaluations and learning: ACF Learning Review 2011' *ALNAP blog*. 18th May. London: ALNAP. www.alnap.org/resource/9964.aspx
- Guerrero et al. (2013) '2012 Learning review'. London: ACF. www.alnap.org/resource/8219.aspx
- Hagen-zanker, J. and Mallett, R. (2013) 'How to do a rigorous, evidence-focused literature review in international development'. London: ODI. www.alnap.org/resource/8675.aspx
- Hallam, A. (2011) 'Harnessing the power of evaluation in humanitarian action'. London: ALNAP/ODI. www.alnap.org/resource/6123.aspx
- Hallam, A. and Bonino, F. (2013) *Using Evaluation for a Change: Insights from humanitarian practitioners*. ALNAP Study. London: ALNAP/ODI. www.alnap.org/resource/8980.aspx
- Hammersley, M. (1990) *Reading ethnographic research: A critical guide*. London: Longman. www.alnap.org/resource/9967.aspx
- Hammersley, M. (2005) 'Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice'. *Evidence Policy* 1(1): 85–100. www.alnap.org/resource/9968.aspx
- HAP (2010) 'The 2010 HAP Standard in Accountability and Quality Management'. Geneva: HAP International. www.alnap.org/resource/8125.aspx
- Harragin, S., and Chol, C. (1998) 'The southern Sudan vulnerability study'. Nairobi: Save the Children and South Sudan Programme. www.alnap.org/resource/9969.aspx

- Hedlund, K. and Knox-Clarke, P. (2011) 'Humanitarian action in drought-related emergencies'. ALNAP Lessons Paper. London: ALNAP/ODI.
www.alnap.org/resource/6156.aspx
- Henry, G. T. and Mark, M. (2003) 'Beyond use: understanding evaluation's influence on attitudes and actions' in *American Journal of Evaluation* 24(3): 293–314.
www.alnap.org/resource/9970.aspx
- Hilhorst, D. (2003) 'Responding to disasters: diversity of bureaucrats, technocrats and local people' in *International Journal of Mass Emergencies and Disasters* 21(3): 37–55.
www.alnap.org/resource/9971.aspx
- Hillier, D. and B. Dempsey (2012) 'A Dangerous Delay: The cost of late response to early warnings in the 2011 drought in the Horn of Africa'. Oxford: Oxfam.
www.alnap.org/resource/6291.aspx
- House, S. (2012) 'Evaluation of Pakistan flood response 2011/12: using Oxfam GB's global humanitarian indicator tool'. Oxford: Oxfam.
www.alnap.org/resource/8366.aspx
- HPG (2006) 'Saving lives through livelihoods: critical gaps in the response to the drought in the Greater Horn of Africa: HPG Briefing Paper'. London: HPG/ODI.
www.alnap.org/resource/8656.aspx
- Humphreys, M. and Weinstein, J. M. (2009) 'Field experiments and the political economy of development' in *Annual Review of Political Science* 12(1): 367–78.
www.alnap.org/resource/9972.aspx
- Huybregts et al (2012) 'The effect of adding ready-to-use supplementary food to a general food distribution on child nutritional status and morbidity: a cluster-randomized controlled trial'. *PLoS medicine* 9(9).
www.alnap.org/resource/9973.aspx
- IASC (2011) 'Humanitarian System-Wide Emergency Activation: definition and procedures'. Geneva: IASC. www.alnap.org/resource/8502.aspx
- IASC (2012a) 'Multi Cluster/Sector Initial Rapid Assessment (MIRA) Provisional version'. Geneva: IASC. www.alnap.org/resource/9974.aspx
- IASC (2012b) 'Operational Guidance for Coordinated Assessments in Humanitarian Crises – Provisional Version'. Geneva: IASC. www.alnap.org/resource/9975.aspx
- ICRC and IFRC (2008) 'Guidelines for assessment in emergencies'. Geneva: ICRC/IFRC. www.alnap.org/resource/6404.aspx
- IFRC (2011) 'Project/programme monitoring and evaluation (M&E) guide'. Geneva: IFRC. www.alnap.org/resource/8542.aspx
- IFRC (2013) 'World disasters report'. Geneva: IFRC.
www.alnap.org/resource/10054.aspx

- Jacobsen, K. and Furst Nichols, S. (2011) 'Developing a Profiling Methodology for Displaced People in Urban Areas'. Boston: Feinstein International Center and Tufts University. www.alnap.org/resource/6306.aspx
- Jaspars, S. (2006) 'From Food Crisis to Fair Trade: Livelihoods analysis, protection and support in Emergencies'. Series no 3. Oxford: Oxfam. www.alnap.org/resource/9842.aspx
- Johnson et al. (2009) 'Research on evaluation use: a review of the empirical literature from 1986 to 2005' in *American Journal of Evaluation* 30(3): 377–410. www.alnap.org/resource/9977.aspx
- Jones, H. (2012) 'Promoting evidence-based decision-making in development agencies'. London: ODI. www.alnap.org/resource/9978.aspx
- Jones, H. and Mendizabal, E. (2010) 'Strengthening learning from research and evaluation: going with the grain'. London: ODI. www.alnap.org/resource/8429.aspx
- Kitson, A. (2002) 'Recognising relationships: reflections on evidence-based practice' in *Nursing Inquiry* 9(3): 179–86. www.alnap.org/resource/9979.aspx
- Knox Clarke, P. (2013a) 'Measuring the effectiveness of humanitarian action'. Presentation at *OCHA roundtable on Humanitarian Effectiveness*, March 2013. www.alnap.org/resource/9980.aspx
- Knox Clarke, P. (2013b) 'Who's in Charge Here?' A literature review of approaches to leadership in humanitarian operations'. London: ALNAP/ODI. www.alnap.org/resource/8640.aspx
- Laybourne, C. and Obrecht, A. (2013) 'Show me the money: generating evidence from and for accountability in humanitarian work', Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March. www.alnap.org/resource/8018.aspx
- LeCompte, M. D. and Goetz, J. P. (1982) 'Problems of Reliability and Validity in Ethnographic Research'. *Review of Educational Research* 52(1): 31–60. www.alnap.org/resource/9981.aspx
- Leeuw, Frans L. (2012) 'Theory Based Evaluation'. www.alnap.org/resource/9982.aspx
- Levine et al. (eds) (2011) 'System failure ? Revisiting the problems of timely response to crises in the Horn of Africa: HPN Network Paper'. London: HPN/ODI. www.alnap.org/resource/6260.aspx
- Macrae, J. (2013) 'Towards an evidence agenda in humanitarian action: reflections on the role of donors in investing in research and innovation in the sector', Presented at, *Evidence in Humanitarian Response: ALNAP 28th Annual Meeting*, Washington, 5–7 March. www.alnap.org/resource/9843.aspx

- Mahmood et al. (2010) 'Guidelines on Data Issues in Humanitarian Crisis Situations'. New York: UNFPA. www.alnap.org/resource/9983.aspx
- Majid, N. (2011) 'Early Warning – Delayed Response?'. Geneva: International Federation of the Red Cross. www.alnap.org/resource/6275.aspx
- Maxwell, D. and Watkins, B. (2003) 'Humanitarian Information Systems and Emergencies in the Greater Horn of Africa: Logical Components and Logical Linkages' in *Disasters* 27(1): 72–90. www.alnap.org/resource/9984.aspx
- Maxwell et al. (2012) 'What Drives Program Choice in Food Security Crises? Examining the Response Analysis Question: Special Edition on Impacts of Innovative Food Assistance Instruments. Boston: Feinstein international. www.alnap.org/resource/9985.aspx
- Maxwell et al. (2013) 'HPN Network Paper: Response analysis and response choice in food security crises: a roadmap', 44(73). London: HPN/ODI. www.alnap.org/resource/9986.aspx
- Mays et al. (2005) 'Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field' in *Journal of Health Services Research & Policy* 10 (Suppl 1): 6–20. www.alnap.org/resource/9987.aspx
- Mazurana et al. (2011) 'Sex and age matter: improving humanitarian response in emergencies'. Boston: Feinstein International Center, Tufts University. www.alnap.org/resource/8144.aspx
- Mills, E. J. (2005) 'Sharing evidence on humanitarian relief' in *British Medical Journal* 331: 1485–86. www.alnap.org/resource/9988.aspx
- Minear, L. (2002) *The Humanitarian Enterprise: Dilemmas and Discoveries*. Bloomfield: Kumarian Press. www.alnap.org/resource/9994.aspx
- Mitchell, G. J. (1999) 'Evidence-based practice: critique and alternative view' in *Nursing Science Quarterly* 12(1): 30–35. www.alnap.org/resource/10000.aspx
- Morra Imas, L. and Rist, R. (2009) 'The Road to Results: designing and conducting effective development evaluations'. Washington: The International Bank for Reconstruction and Development and World Bank. www.alnap.org/resource/8470.aspx
- MSF Vienna Evaluation Unit (2012) 'Evaluation Manual: A handbook for Initiating, Managing and Conducting Evaluations in MSF'. Vienna : MSF. www.alnap.org/resource/10004.aspx
- Nicholson, N. and Desta, S. (2010) 'Evaluation of the Enhanced Livelihoods in Mendera Triangle and Southern Ethiopia 2007–2009'. Washington: USAID and ELMT/ELSE. www.alnap.org/resource/10009.aspx
- OCHA (2007) 'Management response matrix to the Intermediate Review of the Central Emergency Response Fund'. New York: OCHA/CERF. www.alnap.org/resource/10012.aspx

- OCHA (2012) 'Evaluation Initiale Rapide Multi-Cluster sur les inondations au Moyen Chari, Tandjilé, Mayo Kebbi Est – Tchad'. N'Djema: OCHA.
www.alnap.org/resource/10013.aspx
- OCHA (2013) 'Meeting Summary: Consultative Workshop Humanitarian Effectiveness'. New York: OCHA. www.alnap.org/resource/10014.aspx
- OECD-DAC (2002) 'Glossary of key terms in evaluation and results based management'. Paris: OECD/DAC. www.alnap.org/resource/8489.aspx
- Oxfam (n.d.-a) 'How are effectiveness reviews carried out?'. Oxford: Oxfam.
www.alnap.org/resource/8473.aspx
- Oxfam (n.d.-b) 'Monitoring, evaluation, accountability and learning'. Oxford: Oxfam.
www.alnap.org/resource/10015.aspx
- Oxfam (n.d.-c) 'Oxfam GB evaluation guidelines'. Oxford: Oxfam.
www.alnap.org/resource/10016.aspx
- Oxfam (n.d.-d) 'Rough guide to monitoring and evaluation in Oxfam GB'. Oxford: Oxfam. www.alnap.org/resource/10018.aspx
- Oxfam (2004) 'Evaluation of Oxfam GB's Food Aid and Food Security Emergency Intervention in Mauritania'. Oxford: Oxfam and Acacia Consultants Ltd.
www.alnap.org/resource/10017.aspx
- Pantuliano, S. and Wekesa, M. (2008) 'Improving Drought Response in Pastoral Areas of Ethiopia, Somali and Afar Regions and Borena Zone of Oromiya Region'. London: HPG/ODI. www.alnap.org/resource/7510.aspx
- Parker, J. and Sanderson, D. (2013) 'Reviewing the quality of evidence in humanitarian evaluations', Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March.
www.alnap.org/resource/8019.aspx
- Peppiatt et al. (2000) *Buying Power: The Use of Cash Transfers in Emergencies*. London: The British Red Cross. www.alnap.org/resource/10019.aspx
- Piccio, R. (2012) 'Probing the paradox of the RCT craze in international development', *NGO Performance*, 24 May. www.alnap.org/resource/10020.aspx
- Poole, L. and Primrose, J. (2010) 'Southern Sudan: Funding according to need'. Wells: Development Initiatives. www.alnap.org/resource/10021.aspx
- Proudlock et al. (2009) 'Improving humanitarian impact assessment: Bridging theory and practice' in ALNAP Review of humanitarian action in 2009. London: ALNAP/ODI. www.alnap.org/resource/5663.aspx
- Puett, C. and Salpeteur, C. (2013) 'Cost-effectiveness of preventing child morbidity with ready-to-use food in urban Chad'. Presented at, *Evidence and Knowledge in Humanitarian Response: ALNAP 28th Annual Meeting*, Washington, 5–7 March.
www.alnap.org/resource/8009.aspx

- Ramalingam et al. (2009a) 'Innovations in international humanitarian action' in ALNAP Review of humanitarian action in 2009. London: ALNAP/ODI. www.alnap.org/resource/5664.aspx
- Ramalingam et al. (2009b) 'Counting what counts : performance and effectiveness in the humanitarian sector'. in ALNAP Review of humanitarian action in 2009. www.alnap.org/resource/5666.aspx
- Ravallion, M. (2009) 'Evaluation in the practice of development,' in *The World Bank Research Observer* 24(1): 29–53. www.alnap.org/resource/10022.aspx
- Ravallion, M. (2011) 'Knowledgeable bankers? The demand for research in World Bank operations'. World Bank Research Working Paper No.5892. Washington: World Bank. www.alnap.org/resource/10023.aspx
- Redmond et al. (2010) 'A Qualitative and Quantitative Study of the Surgical and Rehabilitation Response to the Earthquake in Haiti' in *Prehospital and Disaster Medicine* 26(06): 449–456 www.alnap.org/resource/7446.aspx
- Robson et al. (2001) 'Guide to evaluating the effectiveness of strategies for preventing work injuries: how to show whether a safety intervention really works'. Cincinnati: CDC and NIOSH. www.alnap.org/resource/10024.aspx
- Rogers, P. J. (2009) 'Learning from the Evidence about Evidence-Based Policy. Strengthening Evidence-Based Policy in the Australian Federation', in *Vol. 1: Roundtable Proceedings*. Canberra: Australian Government Productivity Commission. www.alnap.org/resource/10025.aspx
- Sackett et al. (1996) 'Evidence based Medicine: What it is and What it isn't' in *British Medical Journal* 312(5): 71–2. www.alnap.org/resource/10026.aspx
- Sadler, K. and Catley, A. (eds). (2009) 'Milk Matters: A literature review of pastoralist nutrition and programming responses'. Boston: Feinstein International Center, Tufts University and Save the Children. www.alnap.org/resource/10027.aspx
- Sanderson et al. (2012) Responding to urban disasters'. ALNAP Lessons Paper. London: ALNAP/ODI. www.alnap.org/resource/7772.aspx
- Sandison, P. (2006) 'The utilisation of evaluations' in *ALNAP Review of humanitarian action 2005*. London: ALNAP/ODI. www.alnap.org/resource/5225.aspx
- Schwandt, T. (2009) 'Towards a practical theory of evidence for evaluation', in Donaldson et al. (2009) *What counts as credible evidence in applied research and evaluation practice*. Thousand Oaks: Sage. www.alnap.org/resource/10028.aspx
- Scriven, M. (2009) 'Demythologising Causation and Evidence', in Donaldson et al. (2009) *What counts as credible evidence in applied research and evaluation practice*. Thousand Oaks: Sage. www.alnap.org/resource/10028.aspx

- Seaman et al. (2000) 'The Household Economy Approach: a resource manual for practitioners'. London: Save the Children. www.alnap.org/resource/10029.aspx
- Segone, M. (2009) 'Enhancing evidence-based policy-making through country-led monitoring and evaluation systems' in Segone, M. (ed.) *Country-led monitoring and evaluation systems*. Geneva: UNICEF. www.alnap.org/resource/8567.aspx
- Shaxson, L. (2005) 'Is your evidence robust enough? Questions for policy makers and practitioners' in *Evidence & Policy: A Journal of Research, Debate and Practice* 1(1): 101–12. www.alnap.org/resource/10030.aspx
- Shaxson, L. (2012) 'Expanding our understanding of K*(KT, KE, KTT, KMb, KB, KM, etc.)'. A concept paper emerging from the *K* (Kstar) Initiative conference, 24–27 April*, Hamilton, Canada. www.alnap.org/resource/10031.aspx
- Silverman, W.A. and Sackett, D. L. (1999) *Where's the evidence? Debates in modern science*. Oxford: Oxford University Press. www.alnap.org/resource/10032.aspx
- Slim, H. (2012) 'IASC Real-Time Evaluation of the Humanitarian Response to the Horn of Africa Drought Crisis Synthesis Report'. New York: IASC. www.alnap.org/resource/7761.aspx
- Smith et al. (2011) 'New Technologies in Cash Transfer Programming and Humanitarian Assistance' Oxford: CaLP. www.alnap.org/resource/9634.aspx
- Spence et al. (2003) 'Quality in qualitative evaluation: a framework for assessing research evidence'. London: Government Chief Social Researcher's Office. www.alnap.org/resource/10033.aspx
- Steen et al. (2013) '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/7997.aspx
- Stern et al. (2012) 'Broadening the range of designs and methods for evaluationst'. DFID Working Paper 38. London: DFID. www.alnap.org/resource/8196.aspx
- Tache, B. (2010) 'Participatory Impacts Assessment of Drought Reserve Areas in Guji and Borana Zones', Oromia Region: Save the Children USA. www.alnap.org/resource/9940.aspx
- Telford, J. and J. Cosgrave (2006) 'Joint evaluation of the international response to the Indian Ocean tsunami: Synthesis report'. ALNAP and TEC/ODI. www.alnap.org/resource/3535.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
- UNHCR (2010) 'UNHCR's evaluation policy'. Geneva: UNHCR. www.alnap.org/resource/10034.aspx

- UNICEF (2004) 'UNICEF evaluation report standards'. New York: UNICEF.
www.alnap.org/resource/10035.aspx
- UNICEF (2007) 'Programme policy and procedure manual: programme operations'.
New York: UNICEF. www.alnap.org/resource/10036.aspx
- UNOCHA and Pakistan NDMA (2012) 'Multi-sector Initial Rapid Assessment (MIRA)
Report Pakistan Floods 2012'. Islamabad: Assessment Working Group Pakistan.
www.alnap.org/resource/10038.aspx
- Valid International (2012) 'IASC Real Time Evaluation of the response to the Horn of
Africa Drought: Somalia'. Oxford: Valid International.
www.alnap.org/resource/7506.aspx
- Van de Putte, B. (2000) 'The Utilisation of Evaluation Recommendations in Medecins
Sans Frontieres – Holland: A Study of 10 Evaluation Reports (1997–1999)'.
Amsterdam: MSF-H. www.alnap.org/resource/10040.aspx
- Van de Putte, B. (2001) 'Follow-up to Evaluations of Humanitarian Programmes'.
London: ALNAP/ODI. www.alnap.org/resource/5011.aspx
- Venton et al. (2012) 'The Economics of Early Response and Disaster Resilience:
Lessons from Kenya and Ethiopia: Economics of Resilience Final Report'. London:
DFID and UKaid. www.alnap.org/resource/7758.aspx
- VSF (2009) 'Meat and Milk Voucher project (IMPACT I and II)', Clarke and Fison. Bah
al Gazal, South Sudan: VFS. www.alnap.org/resource/10042.aspx
- Walden, V. (2013) 'Improving quality of humanitarian programmes through the
use of a scoring system: the Humanitarian Indicator Tool', Presented at, *Evidence
and Knowledge in Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7
March. Oxford: Oxfam. www.alnap.org/resource/8024.aspx
- Walker, P. (2013) 'Cracks in the machine: is the humanitarian system fit for
purpose?', Presented at, *Evidence and Knowledge in Humanitarian Action: ALNAP 28th
Annual Meeting*, Washington, 5–7 March. www.alnap.org/resource/8014.aspx
- Willet, B. (2013) 'Mission metrics: An agency wide measurement approach for
understanding emergency response', Presented at *Evidence and Knowledge in
Humanitarian Action: ALNAP 28th Annual Meeting*, Washington, 5–7 March. www.alnap.org/resource/8025.aspx
- WFP (n.d.) 'Monitoring and Evaluation Guidelines. Module: How to design a Results-
Oriented M&E Strategy for EMOPs and PRROs'. Rome: WFP.
www.alnap.org/resource/10045.aspx
- WFP (2005) 'Summary report on WFP follow-up to recommendations'. Rome: WFP.
www.alnap.org/resource/10048.aspx

World Vision Pakistan (2011) 'End of Programme Evaluation Report for DEC-funded WV Relief Program in Sindh (July 2011 – Phase 1)'. Islamabad: Sustainable Solutions Ltd. www.alnap.org/resource/6148.aspx

Young, J. and Court, J. (2004) 'Bridging Research and Policy in International Development: An Analytical and Practical Framework'. London: ODI. www.alnap.org/resource/8430.aspx

Zhang et al. (2002) 'A knowledge management framework for the support of decision making in humanitarian assistance/disaster relief' in *Knowledge and Information Systems* 4(3): 370–85. www.alnap.org/resource/10052.aspx



ALNAP

Overseas Development Institute
203 Blackfriars Road
Londres SE1 8NJ
Royaume-Uni

Tél : +44 (0)20 7922 0388
Fax : +44 (0)20 7922 0399
Site Web : www.alnap.org
Email : alnap@alnap.org

ALNAP aimerait souligner le soutien financier de l'USAID qui a permis de mener à bien cette initiative.

