MONITORING &
EVALUATION

ALNAP

Learning from What We Know

How to improve evaluation synthesis for humanitarian organisations

Neil Dillon

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

www.alnap.org

About the author
Neil Dillon is a Research Fellow with ALNAP.

Acknowledgements
The author would like to thank John Mitchell and Amelie Sundberg for reviewing drafts of this report. Insights and comments from ALNAP Members and participants of the 2018 ALNAP session on evaluation synthesis were also helpful. In particular, James Darcy (independent), Lauren Kejeh (Tearfund) and Deborah McWhinney (WFP) all helped tune the report to the needs and capacities of the ALNAP Membership.

Suggested citation
Dillon, N. (2020) *Learning from What We Know: How to improve evaluation synthesis for humanitarian organisations*. ALNAP Paper. London: ALNAP/ODI

ISBN: 978-1-913526-05-4

Design by Soapbox, www.soapbox.co.uk |

Communications management by Hannah Bass and Maria Gili
Copyediting by Hannah Caddick
Bibliographic editing by Renée Goulet
Cover photo credit: freestocks.org

Contents

1 Introduction

1.1 Why evaluation synthesis is important

The volume of evaluations published by humanitarian actors has increased significantly over the past decade. Annual uploads of evaluation reports to the ALNAP HELP (Humanitarian Evaluation, Learning and Performance) Library increased tenfold between 2007 and 2017. Evaluation policies published in that period by key agencies such as the World Food Programme (WFP), United Nations High Commissioner for Refugees, UK Department for International Development (DFID) and others have all looked to develop and strengthen pre-existing systems, in some cases by committing to increasing evaluation resources as a share of total budget expenditure.

While the publication volume has gone up, the evaluation system still struggles to provide evidence about collective performance of the humanitarian sector. This is partly due to the level of analysis of much of the available evaluation literature, which remains primarily project focused (Darcy and Dillon, 2019). Meanwhile, response-level, country-level and multi-agency evaluations continue to be a relatively small part of the evaluation landscape.

Evaluation synthesis provides one option for building information about response-wide performance from a predominantly project-level evidence base. Syntheses aggregate findings from a defined range of evaluations using a specific analytical framework. They typically seek to provide digestible lessons for senior management by drawing on findings from multiple projects and interventions (ALNAP, 2018). In short, they help us to learn from what we know.

1.2 Current practice and challenges

Some ALNAP Members already conduct evaluation syntheses. The WFP Office of Evaluation 'synthesizes the findings of all centralized evaluations in an annual evaluation report presented to the Board, identifying systemic issues and making overarching recommendations' (WFP, 2016). Likewise, the United Nations Children's Fund (UNICEF) recently published a synthesis of humanitarian evaluations covering 2010–2016 (UNICEF, 2018). And Oxfam worked with the Feinstein International Center to publish a series of systematic reviews under the Humanitarian Evidence Programme, which appraised and synthesised the evidence on the impact or effectiveness of a range of humanitarian interventions and approaches (Oxfam, n.d.).

But many ALNAP Members have expressed a desire for guidance on the methodological choices presented by synthesis work. To this end, the ALNAP Secretariat held a session on evaluation synthesis as part of the 2018 ALNAP M&E Skills-Building Days (ALNAP, 2018). The session briefly covered critical challenges to doing evaluation synthesis in the humanitarian sector and approaches to overcome them. Participants wanted guidance on when and why to conduct synthesis, how to design evaluation synthesis questions well, how to handle variance in the quality and method of evaluations being synthesised and how to ensure synthesis findings are framed meaningfully without generalising beyond their applicable scope. This document has been drafted in response to these requests.

2 When to do an evaluation synthesis

2.1 When is synthesis most useful?

Evaluation synthesis can meet a range of different needs. Most immediately, it can serve a **distillation function**: it takes evidence from an array of evaluation sources and communicates it in a single report, making it easier for decision-makers to access all relevant evidence. The simple act of distilling evidence in this way can improve the usability of evaluative evidence in the decision-making process. The Humanitarian Evidence Programme, for example, was launched in order to 'distill humanitarian evidence and communicate it to key stakeholders in order to enable better decision-making and improve humanitarian policy and practice' (Feinstein International Centre, 2015).

The value of this distillation function increases as organisations produce more evaluation reports on an annual basis. Darcy and Dillon (forthcoming) note that evaluation output across the humanitarian sector increased significantly between 2007 and 2017, with the majority of this increase represented by single-project evaluations, as opposed to larger strategic studies. While these individual project-level evaluations can, in their own right, contribute usefully to the evidence, the use of synthesis methods to distil findings across a sequence of evaluations can further increase its use and increase buy-in from decision-makers (World Bank, 2009: 70). Thus, as the number of relevant studies outstrips absorption capacity, the distillation function becomes more of a powerful tool.

Beyond simply reducing several reports to one, synthesis can also add value by telling us more than any individual evaluation ever can. By abstracting from the details of individual programmes being evaluated, syntheses can generate powerful findings about a cross-cutting issue or theme. We call this the **abstraction function**. As noted by Hallam and Bonino (2013: 78–80):

> Dozens of studies, evaluations and reviews may cover the same issue or theme. Looking at these in their totality, through meta-analysis and evaluation synthesis, can yield far richer evidence and findings.

Both statistical meta-analysis and qualitative approaches can help identify patterns and themes that reach beyond the sum of findings from

the individual evaluations reviewed (Krystalli and Ott, 2015: 3). By looking for effects of similar projects in multiple settings, syntheses can encourage thinking about replicability and generalisability between contexts and help to identify answers to questions like 'what works, in what contexts, and for whom'. For humanitarian evaluations, this abstraction function can be particularly useful in three areas:

1. Higher-level analysis: when looking for findings about themes and issues relating to multiple programme contexts
2. Protracted crisis: when looking for findings across many projects conducted in the same context over time
3. Multi-agency response: when looking for findings that relate to projects funded or implemented by multiple actors

Box 1 The abstraction function in practice

Higher-level analysis:
A good example of this is UNICEF's synthesis of humanitarian evaluations from 2010 to 2016 (UNICEF, 2017). This study synthesised 30 evaluations selected from the total pool of 74 UNICEF humanitarian evaluations between 2010 and 2016. The 30 evaluations were drawn from a wide range of humanitarian crisis responses, sectors and contexts. One of study's main objectives was to identify 'recurrent findings, issues, themes on how UNICEF has performed against its own corporate commitments, standards and stated objectives' (UNICEF, 2017: 54). In that sense, the synthesis asked bigger, more strategic questions than any of the evaluations underlying it. As a result, the evaluators were able to make strategic recommendations for the whole organisation, such as reconsidering UNICEF's Core Commitments to the Child or reviewing how internal UNICEF guidance issued from headquarters is integrated across the decentralised structure of its country operations.

Protracted crisis:
WFP (2018) synthesised four impact evaluations, conducted in the Sahel region, covering protracted food security and nutrition-related responses between 2012 and 2017. The protracted nature of the crisis meant that the synthesis could draw from evaluations covering similar interventions over time. As a result, it was able to abstract from the specific questions addressed by the underlying evaluations and draw conclusions about what was working and what was not working in its food security and nutrition-related programmes. Critically, this allowed the WFP synthesis to provide evidence about the relationship between malnutrition treatment and prevention programmes – an area of comparative data poverty before the study (WFP, 2018: 7).

Multi-agency response:
DEC (2019) synthesised 40 evaluations and wider literature concerning DEC member responses to the 2015 Nepal ('Gorkha') Earthquake. DEC members conduct their own evaluations and reviews to learn lessons at individual agency level. But the DEC synthesis specifically sought to complement rather than duplicate these studies, by focusing attention on the wider context that member agencies were acting in and identifying lessons for future action that apply to all (DEC, 2019: 1).

2.2 What preconditions need to be in place?

Each of the examples in Box 1 highlight the importance of having a sufficient evidence base in place before conducting a synthesis study. UNICEF's synthesis (2017) was based on a total pool of 74 evaluations covering UNICEF's humanitarian response between 2010 and 2016. Each of these evaluations was scored using UNICEF's own quality assurance system and only those considered 'satisfactory' were included in the review. This, alongside the additional exclusion criteria used by the synthesis team, meant that the 30 evaluations reviewed were able to provide a solid basis for synthesis.

In contrast, WFP (2018) was drawn from a much smaller pool of four evaluations but they were conducted using similar methods, asked similar evaluation questions and covered similar topics and contexts. Most importantly, the synthesis sought to ask questions that were fitted to the context in which the underlying evaluations were conducted, namely, WFP's food security and nutrition-related responses in the Sahel. As such, the four evaluations provided consistent and comprehensive coverage of the theme and context under review.

DEC (2019) was able to draw from the numerous studies, reviews and real-time evaluations conducted by its member agencies during the 2015 Nepal Earthquake response. While the multi-agency nature of this synthesis raises questions about comparability of method between evaluations, it did mean that the review team had more sources on which to draw when conducting the synthesis.

The most important lesson here is that the quality and comparability of evidence between evaluations can determine the feasibility of synthesis work. Where evidential quality is patchy, syntheses struggle to draw conclusions with high confidence. Where comparability between studies is low, syntheses struggle to triangulate findings or to assess the strength of the evidence base.

Unfortunately, as noted in Darcy and Dillon (forthcoming), evidential quality and comparability are highly variable in the humanitarian sector. There are a number of issues related to this variability that make synthesis work difficult, including:

- variable definitions of key terms, including interventions and outcomes
- prevalence of study designs that can introduce biases, errors and results

- imitations of project-level data itself
- lack of transparency of methodologies used for evaluations.
  (Krystalli and Emerson, 2015: 3)

This only increases the importance of laying the groundwork for synthesis through the preparation of the evaluation system beforehand. Doing so ensures consistent quality and comparability before synthesis work is commissioned. Here again the example of WFP is worth considering. WFP (2019) provides a synthesis of country portfolio evaluations in Africa from 2016 to 2018, based on a series of evaluations that all have comparable scope and approach. WFP's country portfolio evaluations all seek to assess 'the strategic positioning, decision making, performance and results of all WFP work in a particular country' (WFP, 2019: 3). They all do so by using a similar range of data collection tools, a similar contracting structure and a single quality-assurance system. This greatly reduces – though does not eliminate – the quality variance between the evaluations. It also increases the comparability of findings. As a result, WFP is able to produce country portfolio syntheses of high quality in a timely manner.

It should of course be remembered that data collection methods are ideally determined by the individual evaluation context, for example data availability, access, feasibility of establishing control or comparison groups. So, it would be wrong to expect the same methods to be used across the evaluation portfolio without regard to the specificities of each case. This is especially relevant in the case of multi-agency or multi-context synthesis.

Nevertheless, the evaluation base can still be prepared by building consistent quality control mechanisms and ensuring that methodological clarity is high. When evaluation methods are clear, synthesis teams can interpret findings with greater confidence. For example, a synthesis team may review two evaluations with contradictory assessments of gender-based violence incidence in the same population group. If one evaluation gathered its data through focus group activities, while another used anonymised survey tools, then the reviewers can draw a conclusion with a higher degree of confidence than they might otherwise.

3 How to do an evaluation synthesis well

3.1 Choosing a method

Approaches to evaluation synthesis can vary. In Dillon and Campbell's (2018) review of approaches across the humanitarian sector, they identified three of particular relevance (Box 2).

Box 2 Approaches to evidence synthesis

Systematic reviews are structured approaches that consist of 'mapping out the available evidence, critically appraising the evidence and synthesizing the results' (Krystalli and Ott, 2015). Systematic reviews seek to summarise the totality of the evidence in order to answer high-level questions like 'does this approach or intervention work?'. The process for doing this typically follows a strict and lengthy procedure to minimise author bias, with one recent application to the humanitarian sector identifying 12 steps – from developing the review protocol through to data extraction, synthesis and collation (Krystalli and Ott, 2015). Research questions are typically very tightly defined and focus on measuring the effect of an intervention on a given population group in a given context, usually in comparison to a control. They also typically appraise the literature against pre-defined quality criteria.

Rapid evidence assessments (REAs) 'provide a more structured and rigorous search and quality assessment of the evidence than a literature review but are not as exhaustive as a systematic review' (DFID, 2015). REAs typically seek to answer the question 'what does the evidence tell us about X?' REAs commissioned by DFID typically aim to provide policy-relevant evidential syntheses with a total production timescale of three to six months.

Literature reviews, which are defined here as all reviews of available literature on a given thematic question, seek to synthesise the evidence they present. Literature reviews can vary according to the type of question they aim to answer, the degree to which literature is included or excluded on the basis of structured criteria, and the extent to which

> evidential quality assessments are applied to the literature reviewed and/or integrated into the analysis.
>
> (Dillon and Campbell, 2018: 8)

Evaluation syntheses, as practiced in the humanitarian sector, face a different challenge from systematic reviews. Systematic reviews typically seek to represent the totality of evidence relating to a high-level question, and then select from that wider universe on the basis of clear and distinct exclusion criteria. Such reviews often start with a wider base of hundreds, sometimes thousands, of potential studies, before selecting those that are eligible for review. As such, a great deal of time is invested in the selection process.

Evaluation syntheses, on the other hand, seek rather to summarise findings from a comparatively narrow pool of evaluations commissioned by a single agency in a given period, geographic region or crisis response. This typically reduces the number of reports in contention for review. UNICEF (2017) was potentially the widest-scope evaluation synthesis any single humanitarian agency has conducted, covering seven years of funding worldwide. Yet the total pool of relevant evaluations was only 74, from which the review team excluded 44 before beginning the review.

It is therefore unsurprising that systematic review approaches are rarely used for pure evaluation synthesis. Nevertheless, three of 12 steps identified in Krystalli and Ott (2015) are commonly replicated in high quality evaluation syntheses in the humanitarian sector, namely:

1. definition of the synthesis question
2. quality assessment
3. framing of synthesis findings

The following sections outline considerations for each of these steps, drawing from examples of good practice in the sector.

3.2 Defining the question

Drawing out learning from a series of evaluations can be made harder by the range of different questions the evaluations ask. One clear example of this is the evaluation synthesis conducted by ALNAP for the State of the Humanitarian System report in 2018. The review team looked at 120 evaluations from across the humanitarian sector, including bilateral donors, multilateral agencies, the International Red Cross and Red Crescent Movement, national and international non-governmental organisations and academic and research organisations. The evaluations themselves were typically focused on single-agency responses to specific crises. But the State of the Humanitarian System presents findings at the level of the humanitarian system as a whole. Consequently, there was considerable divergence between the questions asked by the evaluations and the questions asked by the review team.

To address this, the reviewers paid close attention to defining the review question, including defining core terms such as 'humanitarian', 'international humanitarian system', 'effectiveness' and 'impact'. Variance between the evaluations reviewed in their understanding of these terms could have reduced the quality of the synthesis; by investing heavily in their definition early on, the reviewers were able to reduce the potential for misinterpretation during analysis.

Beyond definitions of terms, it is also important for synthesis teams to consider the different audience for a synthesis compared to a single-project evaluation. WFP's country programme syntheses are instructive here. They are indicative of an increased demand for evidence at the level of the executive board, senior management and regional offices. Specifically, they respond to a demand for evidence about the strategic implications of WFP programme results; an additional lens beyond the project-level quantitative data. This is, in essence, an example of the abstraction function described in section 2.2.1. As a result, the questions posed by WFP (2019), for example, are specifically designed to provide findings at this level.

Given these considerations, it is important to ask the following questions when defining the synthesis question:

- How do your audience requirements differ from individual project-level evaluations?
- What are the implications of applying different levels of analysis in the synthesis work (project, country, region, multi-agency)?
- Do your underlying evaluations support findings at the level of the question you are asking?

3.3 Handling quality variance in the evaluation base

UNICEF (2017) benefitted from the presence of a single quality-assurance system having been previously applied to each of the evaluations eligible for review. From the 74 eligible evaluations, only those reaching a score of 'acceptable' evaluation quality were included. For many agencies conducting syntheses of their own evaluations this will also be the case.

But for multi-agency evaluations such as DEC (2019), or for the evaluation synthesis of the *State of the Humanitarian System* report, this is not possible. In such cases, the available options include using an of externally defined set of assessment criteria, such as the Bond evidence principles (Bond, 2013), or designing an entirely new set of criteria tailored specifically to the evaluation synthesis in question.

Dillon and Campbell (2018) explore the pros and cons of using customised criteria designed specifically for the synthesis in question versus using externally defined criteria. In many instances, such as the evaluation synthesis for the 2018 *State of the Humanitarian System* report, there is benefit in developing customised criteria while using externally defined examples as a starting point. Examples of externally defined criteria to consider include the following:

- Christoplos et al. (2017), which looks specifically at the challenges of assessing evidential quality in humanitarian evaluation
- Knox-Clarke and Darcy (2014: 15–17), which provides six criteria specifically aiming to provide a 'tool for humanitarian actors when considering the strength of the evidence that they use'
- Shaxson (2005) and Spencer et al. (2003), which provide evidential-quality assessment tools geared towards policy-makers across a range of fields
- Bond (2013), which proposes five key principles for assessing the quality of evidence produced by NGOs in both the development and humanitarian spheres.

Once an approach has been developed for assessing evaluation quality, the synthesis team will still need to agree how it will be applied to the review. Options include weighting the evidence from high-quality evaluations more highly than that from lower quality evaluations; excluding evidence from low quality evaluations altogether; or using findings from the high-quality evaluations to structure and guide the exploration of findings from other evaluations. Here, again, it is worth consulting reviews of different approaches available in the sector, including Krystalli and Emerson (2015) and Dillon and Campbell (2018).

3.4 Interpreting and framing synthesis findings

Synthesis findings come from a range of different sources and will therefore have varying degrees of evidential support. As a result, it is important that authors are 'clear about what is known and not known, the contexts in which their findings are drawn, and the limitations and gaps in the analysis, so that the readers can make their own assessment' (ibid: 25).

One way to do this is to present an assessment of confidence in each finding, based on clear and explicit criteria. Juillard and Jourdain (2019) conducted a review of lessons learned in earthquake response as part of ALNAP's Lessons Papers series. While not restricted to evaluation synthesis, the review (2019) provides a good example of a synthesis where each finding is provided with a confidence assessment on the basis of clear and consistent criteria. Dillon and Campbell (2018) provide an overview of useful criteria to use when assessing confidence in individual evaluation findings, including considerations around:

- **Evidential quality:** how good were the evaluations that support the finding in question?
- **Consistency across sources:** was the finding disputed by different evaluations under review? If so, what possible explanations might there be for the different perspectives presented, and how does this affect confidence in the finding?
- **Generalisability:** how widely can the finding be applied based on the evaluations supporting it?

(ibid: 25–27)

Beyond providing individual confidence ratings per finding, it is also important for synthesis authors to consider the way in which findings are presented. Given the purpose of synthesis is often to identify trends and patterns across contexts, one option demonstrated by UNHCR (2017) and IASC (2012) is the explicit presentation of findings as 'indicative' of a given issue, accepting directly that it does not constitute 'proof' in the strictest sense. Noting evidence that is indicative of a pattern, outlier results or recurrent issues can in this way be particularly useful for decision-makers (ALNAP, 2018).

4 Conclusion

When done well, evaluation synthesis can provide a valuable information source for humanitarian decision-makers who want to understand what works, for whom and in which contexts. It can distil information in a format that is accessible for management and executive leadership. From the details of a single project it can help to draw out themes and issues spanning multiple programme contexts, actors and over extended periods of time. And it can provide learning that goes beyond the sum of the individual studies reviewed.

For these reasons, evaluation synthesis will be critical in the future of humanitarian learning and accountability. The previous decade has seen a notable expansion of evaluative activity, but the humanitarian system still struggles to provide a significant weight of evidence beyond the project level. Joint, multi-agency and inter-agency evaluations have a big role to play in filling this gap. But evaluation synthesis must also do its part. Synthesis starts from what we already know at the project and the programme level, and agencies can and should be using it more regularly to draw out wider lessons and trends regarding their own performance. Devoting consistent levels of resource to periodic synthesis of evaluative activity will help support strategic decision-making at single agency level and will provide useful input to future collective learning mechanisms.

Bibliography

The following publications can also be accessed via the Humanitarian Evaluation, Learning and Performance (HELP) Library: alnap.org/help-library/evaluation-synthesis-paper-biblio

ALNAP (2018) *ALNAP skills-building days 2018: Evaluation synthesis. Session Notes.* London: ALNAP/ODI (www.alnap.org/help-library/evaluation-synthesis-session-summary-and-tip-sheet)

Bond *(2013) Principles and checklist for assessing the quality of evidence.* London: Bond (www.alnap.org/help-library/principles-and-checklist-for-assessing-the-quality-of-evidence)

Darcy and Dillon (forthcoming) Working title: *Foundations revisited: reflections on current practice in the evaluation of humanitarian action.* London: ALNAP/ODI

Dillon and Campbell (2018) *Lessons Papers: A Methods Note*. London: ALNAP/ODI (www.alnap.org/help-library/lessons-papers-a-methods-note)

Feinstein International Centre (2015) 'The humanitarian evidence program'. [Website]. Accessed 20 November. Massachusetts: Tufts University (www.alnap.org/help-library/the-humanitarian-evidence-program)

Hallam, A. and Bonino, F. (2013) *Using evaluation for a change: Insights from humanitarian practitioners.* ALNAP Study. London: ALNAP/ODI (www.alnap.org/help-library/using-evaluation-for-a-change-insights-from-humanitarian-practitioners)

Juillard, H. and Jourdain, J. (2019) *Responding to earthquakes.* ALNAP Lessons Paper. ALNAP/ODI: London (www.alnap.org/help-library/alnap-lessons-paper-responding-to-earthquakes)

Knox-Clarke, P. and Darcy, J. (2014) *Insufficient evidence? The quality and use of evidence in humanitarian action.* ALNAP Study. London: ALNAP (www.alnap.org/help-library/insufficient-evidence-the-quality-and-use-of-evidence-in-humanitarian-action-alnap-0)

Krystalli, R. and Emerson, L. (2015) *Synthesizing practices of evidence appraisal in the humanitarian field*. Oxford/Massachusetts: Oxfam/Tufts University (www.alnap.org/help-library/synthesizing-practices-of-evidence-appraisal-in-the-humanitarian-field)

Krystalli, R. and Ott, E. (2015) *Evidence synthesis in the humanitarian sector: A humanitarian evidence programme guidance note*. Oxford/Massachusetts: Oxfam/Tufts University (www.alnap.org/help-library/evidence-synthesis-in-the-humanitarian-sector-a-humanitarian-evidence-programme)

Sanderson, D., Patnaik, D. and Osborne, K. (2019) *Nepal earthquakes meta-synthesis*. London: DEC (www.alnap.org/help-library/nepal-earthquakes-appeal-meta-synthesis)

Shaxson, L. (2005) 'Is your evidence robust enough? Questions for policy makers and practitioners'. *Evidence & Policy: A Journal of Research, Debate and Practice* 1(1): 101-112 (www.alnap.org/help-library/is-your-evidence-robust-enough-questions-for-policy-makers-and-practitioners)

Spencer, L., Ritchie, J., Lewis, J. and Dillon, L. (2003) *Quality in qualitative evaluation: A framework for assessing research evidence*. London: Government Chief Social Researcher's Office (www.alnap.org/help-library/quality-in-qualitative-evaluation-a-framework-for-assessing-research-evidence)

UNHCR (2013) *Synthesis report of the joint WFP and UNHCR impact evaluations on the contribution of food assistance to durable solutions in protracted refugee situations*. Geneva: UNHCR (www.alnap.org/help-library/synthesis-report-of-the-joint-wfp-and-unhcr-impact-evaluations-on-the-contribution-of)

UNHCR (2017) *Evaluation synthesis of UNHCR's cash based interventions in Jordan*. Geneva: UNHCR (www.alnap.org/help-library/evaluation-synthesis-of-unhcrs-cash-based-interventions-in-jordan)

UNICEF (2019) *SGBV response, risk mitigation and prevention in humanitarian crises: A synthesis of findings from evaluations of UNHCR operations 2019*. New York: UNICEF (www.alnap.org/help-library/sgbv-response-risk-mitigation-and-prevention-in-humanitarian-crises-a-synthesis-of)

UNICEF (2017) *Towards improved emergency responses: Synthesis of UNICEF evaluations of humanitarian action 2010-2016*. New York: UNICEF (www.alnap.org/help-library/towards-improved-emergency-responses-synthesis-of-unicef-evaluations-of-humanitarian)

UNICEF (2018) *Synthesis of Humanitarian Evaluations 2010-2016 UNICEF Management Response*. New York: UNICEF (www.alnap.org/help-library/synthesis-of-humanitarian-evaluations-2010-2016-unicef-management-response)

UNICEF (2013) *Thematic synthesis report on evaluation of humanitarian action.* New York: UNICEF (www.alnap.org/help-library/thematic-synthesis-report-on-evaluation-of-humanitarian-action)

UNICEF (2012) *IASC real-time evaluation of the humanitarian response to the Horn of Africa drought crisis in Somalia, Ethiopia and Kenya.* New York: UNICEF (www.alnap.org/help-library/iasc-real-time-evaluation-of-the-humanitarian-response-to-the-horn-of-africa-drought)

WFP (2016) *Evaluation policy 2016-2021.* Rome: WFP (www.alnap.org/help-library/evaluation-policy-2016-2021)

WFP (2017a) *Operation evaluations series, regional synthesis 2013-2017: West and Central Africa region.* Rome: WFP (www.alnap.org/help-library/operation-evaluations-series-regional-synthesis-2013-2017-west-and-central-africa)

WFP (2017b) *Annual synthesis of operation evaluations (2016-2017).* Rome: WFP (www.alnap.org/help-library/wfp-annual-synthesis-of-operation-evaluations-2016-2017)

WFP (2018) *Four evaluations of the impact of WFP programmes on nutrition in humanitarian contexts in the Sahel: A synthesis.* Rome: WFP (www.alnap.org/help-library/four-evaluations-of-the-impact-of-wfp-programmes-on-nutrition-in-humanitarian-contexts)

WFP (2019) *Synthesis of country portfolio evaluations in Africa (2016-2018).* Rome: WFP (www.alnap.org/help-library/synthesis-of-country-portfolio-evaluations-in-africa-2016-2018)

World Bank (2009) 'Making smart policy: using impact evaluation for policy making'. *Doing Impact Evaluation* 14 (www.alnap.org/help-library/making-smart-policy-using-impact-evaluation-for-policy-making-case-studies-on)

Related ALNAP publications

- Lessons Paper: A methods note
- Strengthening the Quality of Evidence in Humanitarian Evaluations
- Insufficient Evidence? The quality and use of evidence in humanitarian action
- Foundations Revisited: Reflections on current practice in the evaluation of humanitarian action (forthcoming)

ALNAP
Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ
United Kingdom

alnap@alnap.org