ALNAP Skills Building Days 2018

Evaluation Synthesis | Session Notes

Session Summary Note and Tip Sheet

This Summary Note and Tip Sheet provides a review of the key learning points from the session on evaluation synthesis at the ALNAP Skills-Building Days, on 4-5 June 2018.

It includes the following key learning points:

- Defining evaluation synthesis
- Exploring alternatives to evaluation synthesis
- A summary of key challenges faced
- Tips on how to overcome those challenges

The discussion drew on ALNAP's Methods Note on conducting Lessons Papers (available here) and the Feinstein Centre's work on humanitarian evidence synthesis (available here). Both of these resources are worth consulting for further reflection.

How to get involved

ALNAP considers that quality evaluation synthesis should be a key part of the humanitarian evidence landscape. ALNAP invites further participation from all of those present at the Skills-Building Days, as well as members of our evaluation Community of Practice, and our broader membership base. We would love to speak to you if you have experience of conducting evaluations syntheses for your organisation, or further interest in doing so. We would also be keen gauge member interest in developing ALNAP research or guidance on this topic.

If you would like to get involved or if you have any questions, please don't hesitate to get in touch with Neil Dillon at n.dillon@alnap.org

What is evaluation synthesis?

Evaluation synthesis is an increasingly important part of the humanitarian evidence architecture. Syntheses provide digestible lessons learned for senior management, drawn from a wide range of programmes and interventions. The need for such products in part speaks to an increase in the quantity of evidence being generated by the humanitarian system.

The working definition of evaluation synthesis used in this document is *"an aggregation of findings from a series of evaluations organised through a specific analytical framework"*. There are three important points to note about this definition:

Firstly, the act of aggregation typically involves including some findings and not others. This makes it important for synthesis reports to provide crystal clear criteria for selection and deselection of findings. Without clear explanation of selection criteria, it is hard for the audience to understand the strengths and weaknesses of the analysis or to assess how best to apply the findings.

Secondly, the use of a specific analytical framework allows synthesis reports to go beyond aggregation of evaluation findings. Instead, syntheses often provide new reflections across a region, time period, or theme. Such learning might include the identification of trends, themes and outliers regarding effectiveness, or equally it might identify evidence gaps. But in doing so, synthesis authors are immediately presented with a challenge: how can I generalise findings across different evaluations? How should I present and interpret the evidence claims that the evaluations generate? See section 4 for more on these points.

How do evaluation syntheses differ from other products?

Better Evaluation provide overview information comparing different approaches to evidence synthesis: https://www.betterevaluation.org/plan/synthesize_value. These pages review different theoretical frameworks for undertaking syntheses across evaluations, as well as discussing the challenge of generalising findings. Beyond the frameworks discussed therein, the following three types of activity are worth consideration:

1. SYSTEMATIC REVIEW

What is it?

Systematic reviews are a 'rigorous method to map the evidence base in an [as] unbiased way as possible, and to assess the quality of the evidence and synthesise it' (DFID, 2013: Systematic Reviews in International Development'. A good guide to using systematic reviews is Hagen-Zanker & Mallett (2013), which lays out the key steps and considers some of the challenges to applying this approach to overseas aid.

Typically, systematic reviews ask very specific questions, often following the PICO framework: **P**opulation; **I**ntervention; **C**ontext; **O**utcome. An example of a PICO question might be:

> *What effect do unconditional cash transfers have on improving food security of newly arrived refugees in non-camp urban settings?*

This is a PICO question since it identifies a population (newly arrived refugees), an intervention (unconditional cash transfers), a context (non-camp urban settings), and an outcome (improving food security).

Why is this different from evaluation synthesis?
Thus, systematic reviews ask very specific questions, but they typically ask it of an unbounded evidence base. That is, they seek to represent the universe of evidence on this specific question, rather than asking just what the findings are of a pre-defined series of evaluations from an individual agency over a set time period. In this sense, they are quite different animals from the evaluation syntheses conducted by many of the agencies present at the ALNAP Skills-Building Days.

2. EVIDENCE MAP / EVIDENCE GAP MAP

What is it?
Evidence maps and evidence gap maps present a visual display of what is known about what works in a particular domain. They typically consolidate findings from systematic reviews and impact evaluations, structured around a given outcome or intervention type. This example from 3ie maps the quantity and certainty of evidence about effectiveness for a range of different agricultural insurance mechanisms. 3ie hosts several great examples of evidence gap maps in the international development sector here.

Why is this different from evaluation synthesis?
The primary aim of an evidence map is to describe the evidence density: what types of things do we know a lot about? What types of things do know less about? And thus, where should we direct research efforts?

Evaluation syntheses, as typically practiced in the humanitarian sector, do not take these questions as a primary target. Instead, they aim to gather lessons learned from a given crisis, geographical location or thematic area. Whilst many syntheses do identify evidence gaps as a by-product of this process, it nevertheless remains a secondary priority.

3. META-EVALUATION

What is it?
Strictly speaking, meta-evaluation is the evaluation of evaluation itself. For the original presentation of meta-evaluation in this sense, see Scriven (1969) An introduction to metaevaluation, Educational Products Report, 2, 36- 38. For a more updated breakdown of what meta-evaluation is, and what to bear in mind when doing it, see his 2009 article here.

Why is this different from evaluation synthesis?

Under Scriven's definition, meta-evaluation has a function close to academic peer-review: it provides an assessment of the merits and demerits of an evaluation. Evaluation synthesis is clearly different from this task. Nevertheless, all evaluation syntheses face a challenge regarding the grading of evidential quality from the evaluations that they synthesise. So, as with evidence maps, evaluation syntheses often produce something close to a meta-evaluation as a by-product of their main objective.

What are the critical challenges to doing an evaluation synthesis?

The participants at the ALNAP Skills-Building Day discussed a range of different challenges to evaluation synthesis. The four biggest challenges are summarised below:

1. **Making evaluations fit for synthesis:** the quality of the evaluation synthesis is heavily dependent on having evaluations that provide high quality evidence and do so in a clear and consistent way that supports synthesis across different evaluation reports.

2. **Assessing the quality of evaluations:** finding a clear and consistent approach to assessing evaluation quality that can work for each of the evaluations being synthesised.

3. **Analysis and writing-up findings:** finding appropriate wording to describe findings without overshooting (by going beyond the generalisability limits of the evaluation reports being synthesised) or undershooting (by leaving out all the useful findings present in the original reports).

4. **Timeliness:** how to judge when syntheses should happen: on a periodic basis? or linked to a strategy level decision processes?

How can we overcome these challenges?

Making evaluations fit for synthesis

This involves ensuring evaluation policies have synthesis in mind so that when an evaluation synthesis is done, it can start from a solid base of evaluation reports that facilitate comparison of findings. Things to consider here include ensuring that evaluations have clear statements of:

- Methods used, sampling approaches, and their limitations
- Evaluation scope, evaluation questions and indicators used
- Where findings come from: e.g. affected populations or key informants?
- How many different sources support each finding
- Who conducted the evaluation and who managed and steered it

Assessing the quality of evaluations

The critical issue here is to find a clear and consistent approach to assessing evidential quality across the full set of evaluations under review. This means choosing a clear set of criteria for assessing the quality of evidence within the evaluation reports under review. As outlined in Dillon & Campbell (2018) it is often difficult to take an off-the-shelf criteria list and just apply this to your synthesis. Instead, it is recommended to tailor the criteria to the purpose of your synthesis. Nevertheless, providing clear, transparent and consistent criteria across all evaluations is essential. And the synthesis team will need to ensure that the criteria are applied consistently – whether by asking one expert to review all evaluations, or by working with a small team and comparing their gradings as they go. Alternatively, some agencies get external teams to provide quality assessments once evaluations are completed, which the synthesis team can then use to guide their work from the start of the synthesis project.

There are a number of resources available to guide the selection and design of criteria for assessing evidential quality, including:

- Dillon & Campbell (2018) reviews some of the most common sets of criteria for assessing evidential quality highlights as well as the challenges of applying them to the humanitarian sector
- Christoplos et al. (2017) discusses the challenges of improving evidential quality in humanitarian evaluation
- Knox-Clarke and Darcy (2014) provide six criteria for assessing evidential quality tailored to the nature of evidence in the humanitarian system
- The BOND Evidence Principles also provide a clear set of criteria to use when assessing evidential quality

Analysis and writing-up findings

Unless the synthesis takes a systematic review approach and only draws findings from impact evaluations above an acceptable quality threshold, it is important to refrain from making strong causal statements about effectiveness. Likewise, being aware of the generalisability of the findings reviewed should limit sweeping statements about the findings that apply across the board.

Instead, authors can look for ways to outline evidence that is ‘strongly indicative’ of a given issue, accepting that this does not constitute ‘proof’ in the scientific sense. Nevertheless, noting evidence that is strongly indicative of a pattern or trend, and therefore pointing the way for further research, is still regarded as a very useful result for decision-makers. Thus, the focus of analysis should be on spotting patterns, outlier results, recurrent issues or comparable findings between evaluations.

Lastly, in part because of the difficulty providing findings at this level, it can be helpful to provide a sense of the degree to which high level findings are considered robust by the synthesis authors. This can, for example, be based on an average of the evidential quality scores for the evaluations from which the finding is drawn. Dillon & Campbell (2018) discuss this issue further.

Timeliness

Regarding timeliness, the intended use of evaluation syntheses and the information needs of the target user group, are paramount. Key tips in this regard include:

- Consider doing an approach paper before launching the synthesis. This allows your organisation to discuss and agree the approach and expectations for the synthesis, which helps improve the targeting of the product whilst also building initial interest in the product among the key stakeholders before synthesis work has begun
- Fit the timeframe to the expected uses. It is important to get clarity on the key decision processes for which the synthesis is most likely to be used, e.g., revisions to country strategies, SOPs, annual funding plans. Understanding these processes and the timeline of information needs they create will help to build a timeline for the synthesis that links up well with its key user needs.