

The evolution of humanitarian evaluation

Ian Christoplos

Seminar on “impact evaluations in fragile states”

November 16, 2016



The emergence of humanitarian evaluation

- The Joint Evaluation of Emergency Assistance to Rwanda –where it all started
- Not just about evaluation –also the creation of the humanitarian accountability architecture (ALNAP, SPHERE, HAP, etc.)
- Evaluation perceived as tool for *accountability of the sector* –somewhat different from the “what works” (learning?) agenda driving impact evaluations
- JEEAR was a cathartic process, which revealed many of the challenges we are struggling with today
- Most recommendations have yet to be implemented

After JEEAR

- Concerns shifted from ‘doing’ evaluations to emphasis on evaluation quality and with that holding evaluators (and commissioners?) to account
- Mostly about quality standards (initially through the ALNAP ‘proforma’)
- Next grand debate to emerge was around Kosovo, where the ‘are we doing the right thing’ question was raised (‘why food in a protection crisis’)
- This challenge of addressing the key challenges remains today
- Tsunami Evaluation Coalition –only comprehensive attempt to ‘replicate’ the JEEAR, may have discouraged later attempts (shift to clusters)

What are we learning?

- Evaluation remains a highly sensitive topic, with media attention a significant fear (varies in different countries)
- The struggle for improving quality continues –alongside the RCT debate
- Some commissioners are raising their standards and accepting their own accountability for maintaining quality through monitoring whether their evaluations reflect policies and aims (UN especially)
- Nonetheless, there is still a focus on outputs and limited attention to outcomes (much less impacts) in the vast majority of evaluations

What are we learning?

- Still difficult to reconcile cost/timing/disturbance factors of large evaluations with projectised/timebound humanitarian response
- Some ‘cross-cutting’ issues remain difficult to consistently bring into focus (e.g., localisation, inclusion, context, etc.) –some progress on gender
- ‘Do no harm’ discourse has influenced humanitarian evaluation but has not taken centre-stage
- This is perhaps due to continued discomfort regarding mission creep and humanitarian principles (and perhaps even the results agenda)

What are we learning?

- Protection versus peacebuilding?
- Growing recognition of the need to confront difficult borderline issues, while avoiding using evaluation as a stick to beat humanitarian actors for not addressing political and development failures
- The Agenda for Humanity is so grand that we have little idea where the chips will fall, but will certainly impact on the scope of humanitarian evaluation in the future
- Ever-expanding expectations raise questions about the role of narrowly focused impact evaluations

Quality of evidence in humanitarian evaluation

ALNAP Criteria on quality of evidence	Definition
Accuracy	Whether the evidence is a good reflection of the real situation, and is a 'true' record of the thing being measured.
Representativeness	The degree to which the evidence accurately represents the condition of the whole population, the larger group of interest, the main stakeholders in the intervention, or the diversity that exists in the population.
Relevance	The degree to which a piece of information relates to the proposition that it is intended to prove or disprove.
Generalisability	The degree to which evidence from a specific situation can be generalised beyond a specific response to other situations (particularly important where evidence from one situation is used to create policies applicable to other situations).
Attribution	The extent to which the analysis demonstrates a clear causal linkage between two conditions or events.
Clarity around context and methods	The degree to which it is clear why, how, and for whom evidence has been collected.

Quality of evidence: Five principles

- **Using more robust methodologies for analysis and collection:** using tried and tested approaches from the social sciences for qualitative work, and continuing to explore the possibilities for quantitative and mixed methods approaches
- **Proportionate investment:** ensuring that investments in evidence match the importance of the questions addressed
- **Increased collaboration:** working together to identify key questions; decreasing unnecessary duplication and sharing and challenging results
- **Thinking of the longer term:** collecting consistent data and tracing trajectories
- **Including the knowledge of people affected by crises:** particularly by answering the questions that they need answered

Accuracy

- How to combine 'objective facts' regarding the intervention and also subjective perceptions, while recognising that each are more or less appropriate depending on the evaluative question?
- Has the evaluation considered risks of bias by both informants and evaluators, and proposed possible measures to address these risks?
- When relying on triangulation of evidence, is the evaluation transparent about the weighting and confidence levels of different types and sources of evidence? Where findings highlight contradictory or contrasting perspectives on an issue, is this acknowledged and analysed?

Representativeness

- When a purposive sample is used, is the sample based on an analysis of the different 'stakes' that the respondents have in the intervention?
- Are choices about sampling methods in relation to timeframes, access and available resources made transparent?
- Does the evaluation need an accurate overview of the effects of the intervention in relation to the overall population, or is it more concerned with specific categories of intended aid recipients / programme participants, service users or other stakeholders? Is the evaluation clear about whose voices it intends to 'represent' in the data collected (and whose voices are excluded)?

Relevance

- Is the evaluation explicit about the indicators against which the intervention is being assessed, while also taking a critical stance as to whether those indicators reflect the propositions being interrogated?
- Have steps been taken to proactively encourage the ownership of evaluation findings and recommendations among different sets of decision-makers (including field level staff and service providers) by ensuring that the questions and indicators used provide them with advice and guidance for the decisions they need to make?

Generalisability

- Are the findings of the evaluation meant to be generalisable to other contexts?
- Has the evaluation effort exercised due caution in judging what can be generalised from the findings to other *contexts*, i.e., by making explicit what is unique to a given country, conflict or convergence of risks?
- Has the evaluation effort exercised due caution in judging what can be generalised from the findings to other *interventions*, i.e., by making explicit what is unique to a given type of programme, modality or method?

Attribution and contribution

- Are assumptions about attribution made transparent and clear? Are they backed by acceptable methods? Are any doubts about what can actually be attributed to the intervention explained and alternative explanations of contribution transparently interrogated, where appropriate?
- Does the evaluation acknowledge (and even highlight) the broader contextual factors and other related interventions that together impinge on the likelihood that the intervention contributed to the results claimed?
- Has the evaluation method and process provided space for recognising unintended positive or negative effects of the intervention?

Clarity around context and methods

- Are the basic facts surrounding the evaluation presented, i.e., who commissioned the evaluation, when and why it was commissioned, who conducted the evaluation and what the methods and limitations were?
- Is the evaluation report transparent about: judgements regarding the accuracy of the data; whose perspectives are included (and excluded) in the findings; and what the underlying assumptions are behind the theory of change of the intervention?

Clarity around context and methods

- Have stakeholders (ideally those being targeted or otherwise affected by the intervention) had an opportunity to validate the findings and conclusions of the evaluation?
- Have the selection of methods and decisions about investment of resources for the evaluation reflected the threshold of certainty needed to use and learn from the evaluation?

In sum...

- Discussions on evaluation quality in the humanitarian sector are focused on enhancing quality *within* prevailing constraints
- The vast majority of humanitarian evaluations are inevitably ‘quick and dirty’, but that does not mean acceptance of poor quality evaluations
- The way forward involves reflection on the quality of evidence together with the transparency and rigour with which this evidence is analysed