



## Method Note 2

July 2014

# Improving accuracy in humanitarian evaluations

by Jessica Alexander

In order to draw informed evaluation findings decisions, one of the main tasks for evaluators is ensuring that the data gathered is as accurate as possible. This method note focuses on accuracy in the work of data collection and description conducted as part of an evaluation.<sup>1</sup> The recent ALNAP Study *Insufficient Evidence? The quality and use of evidence in humanitarian action* defines **accuracy** as:

‘whether the evidence is a good reflection of the real situation, and is a ‘true’ record of the thing being measured. Anthropometric measurements that have been conducted correctly, and key informant statements that are a true account of what the informants believe, both pass the test of accuracy. Accuracy can be compromised by the informant or by the observer/researcher, through conscious falsification, unconscious bias, misunderstanding, or the incorrect use of measuring instruments’ (Knox Clarke and Darcy, 2014: 15).

It is important to understand why respondents may not tell the truth, either consciously or subconsciously. If the line of questioning is threatening to them – i.e. it exposes limited skill or knowledge or highlights socially unacceptable behaviour – a risk of what is called desirability bias exists. For example, a mother may not want to tell the truth about how often her children have been fed if she believes that the truth would reflect poorly on her.

Another source of bias is the respondent’s idea of how his/her answer may affect his environment such as a decrease in aid or a change in programme. For example, respondents may exaggerate claims about their conditions and problems if they believe it can further their well-being. When other members of the family or community are present, this may also influence the accuracy of how respondents answer questions. The respondent may be tempted to answer in a way that gives him/her credibility in the eyes of onlookers, rather than giving a truthful reply.

---

<sup>1</sup> This post does not specifically cover the related, but slightly different concept of validity in evaluation, which i.e. refers to validity of findings, interpretations and evaluative judgements made by evaluators. Marian Read has discussed this issue in a recent contribution to the community of practice, see ‘[Are we sure? Veracity in Humanitarian Evaluations](#)’ (Read, 2014). In practitioners and scholarly literature, useful references to ‘validity’ can be found in: Gerring, 2011: Chapters, 4-6 and 7; and Bamberger et al., 2012: Chapter 7.

Other reasons for the provision of false information could stem from faulty memory, misunderstanding of the question or purpose of the interview, or a desire to give answers that the respondent thinks the interviewer wants to hear, called courtesy bias. Even analysis of secondary data is prone to inaccuracy. For example, assessment data from a study in Ethiopia found that government needs assessments data had an economic slant and was more of a 'wish list' than an accurate assessment of needs. In this case, clusters and local NGOs were important in building confidence in government data and its accuracy (Darcy et al., 2013).

The challenge of ensuring accuracy is not only a concern for evaluators, but also, and perhaps even more importantly, is a concern and a challenge for those involved in needs assessment work. As *Insufficient Evidence* notes,

'it is perhaps inevitable that needs assessments – which will form the basis of distributions of limited amounts of scarce and much needed resources – should encounter problems of accuracy. They provide obvious incentives for people to exaggerate the seriousness of a situation to increase the level of support provided. At the same time, governments may underestimate the extent of needs to deny that a crisis is occurring.' (Knox Clarke and Darcy, 2014: 25)

ACAPS' *How Sure Are you?* study (2013) poses questions of accuracy in relation to assessment work and offers guidance to help researchers determine the reliability of the source – i.e. their qualifications, integrity, motives for bias and reputation. There is a whole host of reasons why a respondent may not give accurate information and the paper guides researchers to ask questions such as:

- Does the source have the necessary qualifications or level of understanding to make the claim?
- What relevant experience does the source have?
- Does the source have a reputation and positive track record for accuracy?
- Does the source have a motive for being inaccurate or overly biased?

Based on answers, the paper advises that the source may not be reliable and that researchers should further investigate or seek other sources.

One of the implications of insufficient accuracy and presence of bias in evaluation is that they can lead to subjective analysis and judgement being made by evaluators. The two questions below encapsulate the sort of concerns relating to accuracy and bias in evaluation:

'How can the reader be assured that the findings and conclusions [presented in an evaluation], many of which are drawn from material compiled by one or more researchers often working independently, present a fair and unbiased [description and] interpretation of the situation being studied?

How do we know that another researcher studying the same situation would not have arrived at a different set of conclusions because of a different set of preconceptions?' (Bamberger et al., 2012: 137)

## So what have researchers done about this and how can we attempt to improve accuracy of information collected?

The literature recommends eliminating questions that may provoke untruthfulness. For example, questions should be formulated in a way that reduces the likelihood of their association with negative (or positive) images of the respondents. Alternatively, 'control-questions' should be introduced to detect possible inconsistencies.

Interviewers can also interact with respondents in a way that limits the risk of interference with truthful responses. For example, they should remind respondents that the survey is anonymous and confidential.

Introductions that help remove some of the social stigma from whatever behaviour of inquiry could help remove bias (for example, 'we find that in this area women may not be able to breast feed because of poor dietary conditions. What about you? How often are you able to breast feed?'). To avoid inaccurate responses due to the interviewee remembering incorrectly, O'Neil suggested adding a time reference as part of the question (for example, 'Has your child been ill since the start of the school year?') (O'Neil, 2014). Any suspicion, he suggests, that respondents are not being truthful could be removed by double-checking findings 'with a broader group of stakeholders including project/programme staff. As the staff are emerged in the programme/project on a daily basis, they will have a fair idea as to if your feedback is accurate or not' (ibid).

Finally, researchers can allow respondents to express how much he or she plans on doing something in the future (for example, 'How often do you plan on attending school next week?') allowing the respondent to demonstrate a socially desirable answer.<sup>2</sup>

Medical research deals with the potential for patients and doctors to be subconsciously inaccurate, in particular when testing the therapeutic effect of drugs, by using double blind methodologies. The potential biases in these cases are associated with the risk of doctors detecting greater improvement in patients who have received a drug as opposed to a placebo and with the risk of patients reacting not to the chemical effects of the drug, but to the idea that it is supposed to help them. A double blind approach limits these potential biases.<sup>3</sup>

## Triangulation

Some are of the opinion that '[n]o data-collection method is free from the possibility – even likelihood – of inaccuracy, a main reason triangulation is so important' (Bamberger et al., 2012: 85).

Triangulation has been defined as the 'deliberate attempts to confirm, elaborate, and disconfirm facts and interpretations through multiple data source, multiple methods of data collection, repeated data collection over time, and achieving multiple perspective through involving / consulting multiple researchers' (Bamberger et al., 2012: 137). According to BetterEvaluation, triangulation is a way to validate data by testing

---

<sup>2</sup> <https://www.surveymonkey.com/blog/en/blog/2013/04/24/methodology-measuring-behaviors/>

<sup>3</sup> [http://www.est-translationstudies.org/resources/research\\_issues/Do%20respondents.htm](http://www.est-translationstudies.org/resources/research_issues/Do%20respondents.htm) On the issue of possible bias in survey responses, see for example Stocké and Langfeldt (2004).

‘consistency of findings obtained through different instruments and increases the chance to control, or at least assess, some of the threats or multiple causes influencing our results. Triangulation is not just about validation but about deepening and widening one’s understanding... It is an attempt to map out, or explain more fully, the richness and complexity of human behaviour by studying it from more than one standpoint.’<sup>4</sup>

A recent assessment of the situation in Aleppo in Syria provides an example of rigorous triangulation: enumerators were asked to assess the reliability of key informants, and triangulate informant responses with each other, with secondary sources and with their own observation. Where discrepancies occurred that could not be resolved, the information from the informant was not used (Assessment Working Group for Northern Syria, 2012).

Although most evaluation mention triangulation as one key tactic to ensure accuracy, they rarely describe how triangulation was done in practice; what that meant for the findings; how much weight we can put on them; and whether we can draw reasonable conclusions based on them. Poor triangulation methods can have considerable implications for the accuracy of the data. As noted in *How Sure Are You?*

‘contradictions and inconsistencies are often discarded or categorised as untrue, when their analysis could be of great value and represent variability in the observed setting. When using qualitative research to study a complex phenomenon, social researchers suggest that convergence of data should not be expected when performing triangulation. Instead of a tool for validation, the triangulation process should offer different perspectives and give access to different versions of the phenomenon under scrutiny.’ (ACAPS, 2013: 17)

They go on to further advise that if triangulation does not reduce the number of alternate explanations for the findings, that researchers should seek a peer review process whereby colleagues review the findings and provide an outside perspective and expertise (ACAPS, 2013). Similarly, O’Neil suggested to

‘double-check any large variations with the project staff and other persons being interviewed. Sometimes you will find that there are separate distinct perceptions of the same issue/ programme - or there is a dominant view with a minority view also. Evaluators can report such differences in their reports.’ (O’Neil, 2014)

## Measuring against a constant

Save the Children’s *Household Economy Approach* provides advice for obtaining accurate information from families on household income and expenditure. It suggests that the interviewer keeps a running tally of information reported so as to pick up inconsistencies. O’Neil points out that

‘evaluators are often parachuted into field situations where their schedule is pre-populated with persons to be interviewed set up by the very organisation under evaluation...which can be an issue.’ (O’Neil, 2014)

---

<sup>4</sup> <http://betterevaluation.org/evaluation-options/triangulation>

Evaluators could consider having ready their own breakdown of the categories and groups of informants/people/profiles that they ideally would like to cover during the field visit and then compare this with the schedule suggested by the 'host' agency in the field.

'In terms of accuracy, this just gives you more confidence that you are getting a broader range of views and not simply hearing from the "converted".' (O'Neil, 2014)

Evaluators and interviewers should also balance the budget at the time of the interview to determine if people are deliberately withholding information, such as income due to illegal activities. Seaman recommends reconciling the faulty information given during the interview so that if it does not add up and it is clear that the responses are incomplete, that information should be discarded. Authors of *The Household Economy Approach* also advise measuring the amounts reported by reasonable averages for people and if it is reported much higher or lower, asking for an explanation (Seaman et al., 2000).

Sometimes areas of inquiry can be verified by evaluators themselves. For example, interviewers can walk the distance that respondents have to get water in order to verify the distance reported. These alternative methods are useful when measuring against a known entity. These can be useful in considering the likely margin of error in data and its significance for the conclusions being drawn or the calculations being made.

'The nature of the information you are gathering about how people live in a rural society makes it possible to check for consistency. On the one hand, there is a finite and relatively small number of economic options available to households, so you will already know the broad parameters of the investigation. You will be looking at how people combine a variety of known options to survive. Cross-checking, then, is simply a way of confirming that all the pieces of information you obtain fit together in a coherent way, and that the information set enables the household to survive in a normal year.' (Seaman et al., 2000)

## Neighbourhood method

Another method used to get around potential bias when asking sensitive questions is called the 'neighbourhood method'. This method was used by researchers who had difficulty gathering accurate data about Gender-Based Violence (GBV) in certain settings, preventing humanitarian and human rights actors from identifying effective prevention interventions. Columbia University's Child Protection in Crisis (CPC) Network has used the 'neighbourhood method' to gather population-based data on rape and other forms of GBV in five countries recovering from conflict. The method taps informal sources of information and operates on the assumption that people know when their neighbours have been raped or beaten and that ordinary individuals can serve as sources of information about other people in their neighbourhoods. Specifically, the method surveys women about their own experience, their sisters' experiences, and their neighbours' experiences.<sup>5</sup>

---

5 <http://www.cpcnetwork.org/learning-details.php?ID=1> For an example see Potts, Mayers and Roberts (2011)

## Aggregation of answers

Other methods rely on repetition of asking the same question over time and aggregating answers. Accuracy is concluded if the analysis shows that the aggregation of the answers is consistent. For example, Oxfam has used the 'visual analog scale' with communities in Ethiopia as part of their Drought Early Warning System which asks multiple people the same questions/indicators about malaria and milk production, plantings and harvests, livestock deaths and births, but then compares it with anecdotal comments gleaned from the villagers as well as statistics gathered from district markets and health posts. Using multiple sources of information provides them with enough evidence to be confident about their early warning system (Chan, 2014).

## References

**ACAPS (2013)** *How sure are you? Judging quality and usability of data collected during rapid needs assessments* (Technical Brief). Geneva: Assessment Capacities Project.

**Assessment Working Group for Northern Syria (2012)** *Joint Rapid Assessment of Northern Syria: Aleppo City Assessment - March 2013*.

**Bamberger, M., Rugh, J., Mabry, L. (2012)** *Real World Evaluation – Working under budget, time, data and political constraints* (2nd edition). SAGE Publications.

**Chan, J. (2014)** 'Oxfam America's community-based Drought Early Warning Surveillance approach', Presented at *Engagement of crisis-affected people in humanitarian action*, ALNAP 29th Annual Meeting, Addis Ababa, 11-13 March.

**Darcy, J, Stobaugh, H., Walker, P. and D. Maxwell (2013)** *The Use of Evidence in Humanitarian Decision Making*. ACAPS Operational Learning Paper. Boston: Feinstein International Center and Tufts University.

**Gerring, J. (2011)** *Social Science Methodology A Unified Framework* (2nd edition). Cambridge: Cambridge University Press.

**Knox Clarke, P. and Darcy, J. (2014)** *Insufficient evidence? The quality and use of evidence in humanitarian action*. ALNAP Study. London: ALNAP/ODI.

**O'Neil, G. (2014)** *Discussing accuracy in humanitarian evaluations – Round-up for final comments* [ALNAP Humanitarian Evaluation Community of Practice]. 14 May. Available at <https://partnerplatform.org/alnap/humanitarian-evaluation/discussions/61469506>.

**Potts, A., Mayers, K., and Roberts, L. (2011)** 'Measuring human rights violations in a conflict-affected country: results from a nationwide cluster survey in Central African Republic', in *Conflict and Health* 2011, 5:4, 14.

**Read, M. (2014)** *Are we sure? Veracity in Humanitarian Evaluations* [ALNAP Humanitarian Evaluation Community of Practice]. 7 April. Available at <https://partnerplatform.org/alnap/humanitarian-evaluation/discussions/7c1ca59b>.

**Seaman, J., Clarke, P., Boudreau, T. and J. Holt (2000)** *The Household Economy Approach: A Resource Manual for Practitioners*. London: Save the Children.

**Stocké, V. and Langfeldt, B. (2004)** 'Effects of Survey Experience on Respondents' Attitudes Towards Surveys', in *Bulletin of Sociological Methodology*, Vol. 81, 5-32.

This Method Note is the result of a discussion on 'Discussing accuracy in humanitarian evaluations – Round-up for final comments', which took place among members of the ALNAP Humanitarian Evaluation Community of Practice (CoP) between April and May 2014.

The CoP discussion presented in this brief note has been enriched with references to evaluation literature and guidance provided by Jessica Alexander. Whenever possible, these notes highlight the experience from ALNAP Members in dealing with specific evidential challenges in humanitarian evaluation practice.

ALNAP's Humanitarian Evaluation CoP provides a space to:

- Discuss issues related to humanitarian evaluation
- Share resources and events
- Consult peers

**ACCESS THE CoP HERE:**

<https://partnerplatform.org/humanitarian-evaluation>