



SEADS Evidence Checklist

A tool for assessing the strength of evidence in evaluations, reviews, and impact assessments

April 2021

Contents

Introduction.....	2
SEADS Evidence Checklist.....	2
Indicator 1: Livelihoods context is clearly understood and described.....	3
Indicator 2: Livelihoods impacts are measured and analyzed.....	3
Indicator 3: Evaluation design is aligned to intervention objectives.....	4
Indicator 4: Design, methods, biases, and limitations are clearly described.....	4
Indicator 5: Different interventions are compared, where possible.....	5
Indicator 6: Mixed methods and triangulation are used.....	5
Indicator 7: Measurements and indicators are relevant.....	6
Indicator 8: Sampling is appropriate relative to time and cost.....	6
Indicator 9: Social equity component is considered.....	7
Indicator 10: Plausibility statements are included.....	7
Indicator 11: Attribution is analyzed.....	8
Indicator 12: Findings flow clearly to the conclusions.....	8

Standards for Supporting Agricultural Livelihoods in Emergencies

<https://seads-standards.org>
coordinator@seads-standards.org

Introduction

SEADS follows an evidence-based approach to developing standards and guidelines, and this process includes reviews of documents that provide evidence on the impact and cost effectiveness of agriculture interventions undertaken in the context of different kinds of crises including, rapid onset, slow onset, complex, and protracted crises. In particular, SEADS seeks evidence of interventions that are shown to be timely, relevant and have positive impacts on livelihoods, markets, systems, and services

This Evidence Checklist provides guidance on how to assess the quality of evidence in documents covering smallholder agriculture-related interventions before, during, and after emergencies. The Checklist is a point of reference for the SEADS Steering Group, consultants recruited by SEADS to produce evidence reviews, and other users. It is based on guidelines for evaluation produced by DFID¹ and USAID²; experiences with developing the evidence database of the Livestock Emergency Guidelines and Standards (LEGS) from 2006 to 2019; commonly-used criteria for peer review of scientific journals papers.

The Evidence Checklist is working document, which can be revised regularly based on feedback from users, and as more studies and evaluations become available.

SEADS Evidence Checklist

Indicators:



¹ DFID (2014). Assessing the Strength of Evidence. How-To-Note. March 2014
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf

² USAID (2016). Evaluation Policy, USAID, Washington DC
<https://www.usaid.gov/sites/default/files/documents/1870/USAIDEvaluationPolicy.pdf>

Standards for Supporting Agricultural Livelihoods in Emergencies

<https://seads-standards.org>
coordinator@seads-standards.org

Indicator 1: Livelihoods context is clearly understood and described

Does the report reflect an understanding of local livelihoods context?

It is widely recognized that good project design and good research design requires an understanding of local context. This is why research often involves an initial qualitative “exploratory” phase to document local histories, conditions and issues, and then identify research questions that are relevant to context. Similarly, many successful projects are based on an understanding of what people in target areas are already doing and why, as well as the wider social, political, and institutional arrangements.

Indicator 2: Livelihoods impacts are measured and analyzed

Does the report measure livelihood impacts at household level, or only project activity and implementation?

A common limitation of project M&E systems is that they focus heavily, or exclusively, on the measurement of project activity, with limited attention to the impacts of activities at household level. One outcome is that over time, project staff might perceive project implementation as automatically resulting in impact—even when these impacts are never actually measured. Similarly, evaluations can involve the collation of monitoring data and summaries of project activities, and these are sometimes used as measures of performance. However, this is a relatively weak approach if used independently of impact measurements.

Indicator 2 is based on the notion that ultimately we want to understand how an intervention has supported livelihoods, if at all. Typically, this means some kind of measurement of livelihoods assets, such as indicators of household financial assets or human assets.

Box 1: Examples of livelihoods impact indicators in emergency agriculture projects

- Household consumption of own-produced agriculture produce, such as cereals, vegetables, or fruit
- Analysis of the nutritional value of own-produced foods against recommended daily allowances of nutrients
- Income derived from the sale of agricultural produce or value-added produce sales (eg. flour from grain, peanut butter from peanuts)
- Proportion of income from the sale of agricultural produce relative to total income
- Specific uses of income from the sale of agricultural products e.g. to buy healthcare, pay school fees; buy other foods; invest in further agriculture production; cash savings.
- Diversification of income streams through yield-dependent food related enterprises
- Application of knowledge acquired through training, leading to improved consumption or income from agricultural production

Note – improved agriculture production is a weak livelihoods impact indicator, unless it is combined with measures of home consumption of produce, or income derived from produce sales.

A livelihoods-based approach in emergencies can also include support to the systems and services that are needed to support livelihoods during or after the emergency. In the case of agriculture interventions, these systems and services include markets, financial services, private sector input suppliers, and government extension or plant pest surveillance services. Therefore, one aspect of understanding

livelihoods impacts is assessing the direct or indirect impact of an intervention on systems and services of relevance to agriculture.

Indicator 3: Evaluation design is aligned to intervention objectives

Is the evaluation or study designed to cover each of the intervention's objectives, with the evaluation questions aligned to these objectives? Is the report structured accordingly?

The questions in an evaluation or impact assessment should relate clearly to some or all of the intervention's objectives. For example, if an intervention aims to improve crop production, at least one of the evaluation questions should aim to measure crop production and the uses of the crops or crop products.

Sometimes the objectives of an intervention change but these changes are not properly documented. Therefore, in the minds of project managers or coordinators the intervention is aiming for one of more objectives that do not fit the original planning document. In these cases, it is common for this mismatch to be carried over into evaluation questions.

Indicator 4: Design, methods, biases, and limitations are clearly described

Does the report include a logical a clear description of the evaluation or study design and methods? Are the limitations of the work explained, with implications in terms of the quality of evidence and the report's conclusions?

In a well-designed project, each objective has a set of activities that relate directly to the objective. This is part of the project's causal logic, logical framework, or theory of change. A similar approach applies to evaluations and impact assessments, and each key question should be addressed using one or more methods that relate specifically to that question. A common weakness in evaluation reports is that methods are described, but it is unclear which methods were used to answer each question. This can be important when results are presented, especially if it is also unclear how the results were derived.

Every evaluation, impact assessment or study has design and methodological limitations. If the authors of a report explain these limitations, and if and how they were handled, it helps the reader to assess whether any major weaknesses were properly addressed. Central to this assessment is how the design of the work addressed the two main types of **bias** – sampling errors and non-sampling errors. Sampling errors arise due to inappropriate sampling of villages or households (or other units), or an inappropriate sample size; this is covered under the "Sampling is appropriate relative to time and cost" indicator.

Non-sampling errors include problems with communication; language and trust; and the relationship between an interviewer, researcher or facilitator, and an informant or informant group. Non-sampling errors are often overlooked in evaluation reports or studies, but can have a substantial impact on findings. Some common non-sampling errors are:

- Language – key words are wrongly translated or explained in interviews or questionnaires, or when using participatory methods. Examples of potentially confusing words, difficult to translate, are "livelihoods," "resilience," and "empowerment." Various technical words can also be difficult to translate e.g. the scientific English names of crop parasites or diseases.
- Trust – the type of information provided by an informant partly depends on whether or not they trust the interviewer, and if they understand how the information they provide is being used. Mistrust can lead to under-reporting e.g. of financial or physical assets.
- Sensitive questions – questions on sensitive issues may lead to inaccurate answers, even if the respondent trusts the interviewer.
- Survey fatigue – occurs when people are exposed to numerous surveys and expected to spend many hours answering questions; it leads to boredom and inaccurate responses.

Standards for Supporting Agricultural Livelihoods in Emergencies

<https://seads-standards.org>
coordinator@seads-standards.org

- Expectations – are a particular issue if a project provides free or subsidized inputs. Informants may exaggerate the benefits of this support in the hope that the support will continue, or, understate the benefits in the hope that the support will not only continue, but also increase in amount or coverage.

One of the main approaches for reducing non-sampling errors is to pre-test a questionnaire or set of participatory methods, and then adjust the method as needed. If pre-testing is not mentioned in a report, along with approaches for ensuring accurate translation, the reader has to assume that important errors in the methodology were overlooked.

Indicator 5: Different interventions are compared where possible

Does the report use an evaluation or research design that includes comparison of the performance, impacts and cost-effectiveness of different interventions?

For many types of interventions, there is no “gold standard” or recognized national or international target to be reached. If so, end-of-project targets are often arbitrary and estimated. In terms of research or evaluation design, one of the most convincing approaches to produce strong evidence involves comparison; various comparative designs are available. Randomized case control trials are an example of a comparative approach, where “treatment” groups receiving project support are compared with “control” groups, not receiving project support. However, this approach is operationally difficult, and possibly unethical, in humanitarian contexts. Other, more practical approaches involving comparison include comparison of different interventions with similar objectives. One or more of the interventions is attributed to a project, whereas other “interventions” are non-project activities.

Box 2: Improving farmer knowledge and practice

Assume that an impact evaluation aims to assess the impact of farmer field schools (FFS). The evaluation could compare the knowledge and practice of FFS participants with farmers who received information or learning support from other sources e.g. public radio, social media, agriculture input suppliers, local faith-based organizations, or market actors. The evaluation could also compare the cost effectiveness of each source.

Indicator 6: Mixed methods and triangulation are used

Does the report use the right mix of qualitative and quantitative methods, and/or a mix of methods to allow triangulation? Are inconsistencies in results from different methods explored and explained?

Much of the debate around the evaluation of humanitarian aid focuses on methodologies and the relative merits of qualitative and quantitative approaches.

- Quantitative surveys are often seen as producing the best quality evidence, but often suffer from limited understanding of context and therefore, flawed designs or unexplained results. Findings can also be statistically significant, but have limited livelihoods significance. In part, these weaknesses can be overcome by better qualitative analysis i.e. a mixed method approach.
- Qualitative methods can produce rich contextual analysis but can suffer from limited coverage of project areas or are not representative; results can be difficult to summarize and the quality of the work depends heavily on the experience of the evaluator(s).
- Validity of findings is enhanced through triangulation i.e. crosschecking results derived from one method with results derived from another method. The methods can be varied, and include quantitative, qualitative, and participatory methods. *In difficult operational contexts, and where time and money are limited, triangulation using mixed methods often provides an acceptable level of evidence.*

Standards for Supporting Agricultural Livelihoods in Emergencies

<https://seads-standards.org>
coordinator@seads-standards.org

The main guidance in terms of evidence quality, is that a study or evaluation that uses only a single method (e.g. a quantitative survey, using a questionnaire for data collection) runs a higher risk of producing results of limited value relative to a mixed-method approach. Practical considerations such as the time and funds assigned to an evaluation often prevent the use of well-designed quantitative surveys. The use of plausibility statements also enhances evidence quality (see Indicator 10).

Reference

Victora, C.G., Habicht, J-P. and Bryce, J. (2004). Evidence-Based Public Health: Moving beyond Randomized Trails. *American Journal of Public Health* 94/3, 400-405.
<https://www.ncbi.nlm.nih.gov/pubmed/14998803>

Indicator 7: Measurements and indicators are relevant

Did the work measure relevant indicators in the right way? Is the analysis meaningful? If quantitative approaches are used, do the results have both statistical significance and livelihoods significance?

Indicator 1 of the checklist, dealing with livelihoods context, is important because a good understanding of context guides the selection of questions, measurements, and indicators in an evaluation or study. The use of inappropriate indicators can lead to results that have statistical significance but limited, or no, livelihoods significance. For example, assume that an impact evaluation compares sorghum yields in intervention vs. non-intervention households, and uses yield as the main impact indicator. The evaluation might find a statistically significant higher yield in project households and therefore, conclude that the project was a success. However, the project did not measure income derived from sorghum sales, and so overlooked substantial post-harvest losses, and local market failures. When these issues were considered, the intervention had limited (non-significant) livelihoods impacts.

A further consideration is that projects commonly report their performance using absolute measures of activity, but with no reference to a population. This is often misleading. For example, assume that as part of a drought response program in Kenya, an NGO reports providing seeds and tools to 5,000 households. At first sight, this seems like an impressive activity. However, what if 50,000 households required assistance? In this situation, the intervention covered only 10 percent of need.

Indicator 8: Sampling is appropriate relative to time and cost

What was the sampling method and how was the sample size determined? Was representative sampling needed and if not, what approach was used and why? Are sampling issues related to the agricultural context understood?

Conventionally, three main types of sampling are used:

- Random sampling – in which the chance of a village, household, or individual being selected is equal to the chance of any other village, household, or individual being selected.
- Purposive sampling – where project staff or evaluators make judgements about which areas and individuals best represent the project under examination.
- Convenience sampling – e.g. where villages are selected based on security or logistical issues.

All three types of sampling are affected by limitations in time and budget.

Of these three types of sampling, random sampling is usually regarded as most representative of a population and therefore, more objective; it can be used with quantitative, qualitative, or participatory data collection methods. Purposive sampling can work well, but only when there is a genuine commitment to the evaluation or study by the implementing agency, and for example, a willingness to identify areas where the project is thought to have had different levels of implementation and impact. In contrast,

Standards for Supporting Agricultural Livelihoods in Emergencies

<https://seads-standards.org>
coordinator@seads-standards.org

purposive sampling can lead to evaluators being directed to areas with relatively strong project activities and impacts. Convenience sampling is sometimes the only option, but is unlikely to be representative of the whole project area.

It is beyond the scope of the Evidence Checklist to provide detailed guidance on the various sampling methods, and methods for calculating sample size, but quantitative surveys or studies using random sampling do not necessarily lead to useful results. Many of the limitations relate to misunderstanding the local context (see Indicator 1).

Indicator 9: Social equity component is considered

Did the evaluation include specific questions on the project's impacts on women or girls? Was sufficient time and effort dedicated to involving women or girls in the evaluation or study? Were the findings presented by gender, and if not, why not?

It is good practice in evaluations and impact assessments to recognize that women and girls are often among the most vulnerable people in communities, and may benefit from humanitarian assistance in different ways to men and boys. Benefits may be skewed towards men and boys, and women and girls may be excluded or experience negative impacts. It is also recognized that in mixed gender focus group discussions or similar events, the discussion is usually dominated by men. Not only are specific sessions and interviews often needed with women without the presence of men, but skilled female evaluators, facilitators, and data collectors are needed.

When reviewing an evaluation or study, the following issues can be considered:

- Does the evaluation or study include specific questions related to gender, and the need to assess the project from a gender perspective?
- Does the design and methodology for the work include specific details on how the perspectives of women and girls were understood? Were the methods appropriate given the cultural context, and did they avoid bias or interference from men?
- Were the results presented by gender, and if not, why not? Were differences between female and male informants explained?
- Were gender specific conclusions and recommendations needed? Were these presented and justified?

Indicator 10: Plausibility statements are included

Did the evaluation include plausibility statements? Was the project's causal logic correct? Were the reported impacts or outcomes plausible relative to the inputs?

Regardless of whether quantitative, qualitative, or mixed methods are used for an evaluation or impact assessment, the results need to be technically plausible. This means that the measured impacts or outcomes of an intervention need to be consistent with the type, amount, timing and quality of inputs. In other words, both the causal logic of the intervention and the findings of an evaluation, need to be plausible. Furthermore, the findings need to be plausible, irrespective of any statistical findings.

Determining whether or not an intervention has a strong causal logic or plausibility needs a good understanding of both local context (Indicator 1), and the technical aspects of the intervention design. Box 3 examines the plausibility of a cash-for-work project. Note that an intervention can have a strong causal logic, but still fail to deliver the expected impacts. This happens if there are likely problems with implementation and activities did not take place as expected.

Box 3: Economic plausibility in cash-for-work project

A cash-for-work project targets vulnerable farming households and aims to support food security, build assets, and protect assets. It is assumed that most households would invest extra cash in some agricultural assets.

A review of the project examined household economies, the prices of local foods, and the food requirements of households. A simple analysis was then done to estimate how the cash transfers would contribute to household economies for more vulnerable households. This analysis showed that the cash income would make a substantial contribution to the basic food needs of households, but also, very little cash would be left over to invest in productive assets. In other words, the project was likely to support food security, but not help households to build assets - the asset-building objective was implausible. It was further concluded that the asset protection objective of the project was also not plausible, because the most vulnerable households had no productive assets to protect, and very few physical assets.

Indicator 11: Attribution is analyzed

Does the report summarize other important projects in the area, run by aid agencies or government? Are these other projects considered and if not, why not? How does the evaluation or study demonstrate that project impacts are specific to the project in question, and not due to other projects, or combined activities?

Evaluations and reviews of projects often assume that positive or negative impacts are due solely to the project in question. Typically, projects are assessed with limited attention to other projects or programs which are recent or ongoing, and which complicate the assessment of attribution. A situation can also be complex because projects in other sectors can have an important influence on the performance of the project being evaluated. In humanitarian contexts, multiple agencies and projects can be present, with weak coordination, and it might be difficult to know “who is doing what and where.” In this situation, analysis of attribution is both necessary, but also, challenging. Important economic and social changes in communities can also arise which have nothing to do with government or aid projects, but happen through local innovations and adaptations, or private sector activity.

Despite the difficulties of measuring attribution quantitatively, a qualitative analysis is often possible. This might include a brief listing of other important projects in the area, and how they might contribute to livelihoods, health or nutrition. “Non-project” factors can also be important, such as how rainfall patterns might affect crop production.

Indicator 12: Findings flow clearly to the conclusions

Do the findings relate clearly to the key questions of the evaluation and therefore, the intervention objectives? Do the findings also relate clearly to methodology, so it is clear which method produced the findings, with which informants?

In a similar way that a report’s key questions should relate directly to the intervention’s objectives, the report’s findings should relate to, and be structured around, the key questions. Plus, it should also be clear which method produced each set of findings, and the number and type of informants who provided the information.

With quantitative evaluation approaches, results are often presented in the form of statistics. Here, the risk is that some readers will not fully understand the meaning of these statistics; if the statistics are not understood, they may not be trusted. As an analysis becomes more complex, the statistics also become more complex, leading to a report that is only really understood by statisticians, but not practitioners. This problem can be partly overcome if explanatory notes are included which guide readers in their interpretation of the results.

In the conclusions or recommendations sections of reports, there should be clear reference back to a set of results, and the conclusions should not “appear from nowhere.” Conclusions or recommendations that are not tied to specific findings may indicate biases during the evaluation, or information that was collected but not presented in the report’s findings.