

Consider important aspects of the evaluation

Evaluations are designed to answer the [Key Evaluation Questions](#). Different types of questions need different methods and designs to answer them.

In evaluations there are four main types of questions:

Descriptive questions ask about what has happened or how things are – for example:

- What were the resources used by the program directly and indirectly?
- What activities occurred?
- What changes were observed in conditions or in the participants?

Causal questions ask about what has contributed to changes that have been observed – for example:

- What produced the outcomes and impacts?
- What was the contribution of the program to producing the changes that were observed?
- What other factors or programs contributed to the observed changes?

Evaluative questions ask about whether an intervention can be considered a success, an improvement or the best option and require a combination of explicit values as well as evidence – for example:

- In what ways and for whom was the program successful?
- Did the program provide Value for Money, taking into account all the costs incurred (not only the direct funding) and any negative outcomes.

Action questions ask about what should be done to respond to evaluation findings – for example:

- What changes should be made to address problems that have been identified?
- What should be retained or added to reinforce existing strengths?
- Should the program be refunded?

Key Evaluation Questions often contain more than one type of questions – for example to answer the KEQ “How effective has the program been?” requires answering:

Descriptive questions – What changes have occurred?

Causal questions – What contribution did the intervention make to these changes?

Evaluative questions – How valuable were the changes in terms of the stated goals – taking into account the types of changes, the level of change and the distribution of changes.

Check the adequacy of the design by disaggregating each KEQ into the different types of questions and then checking them against the following points.

(i) Checking the adequacy of the design for descriptive questions

The design should make it clear how descriptive questions will be answered. These descriptive questions might relate to:

- Inputs – materials, staff
- Processes – implementation, research projects

- Outputs – eg research publications
- Outcomes – eg changes in policy on the basis of research
- Impacts – eg improvements in agricultural production

It can be helpful to set this out in a table that shows how data will be collected and analysed to answer these descriptive questions.

Descriptive question	Existing data that can be used	Additional data collection/retrieval	Sampling Analysis
What has been the level of resources used for the program?			
Who has participated in the program?			
What changes have occurred in terms of [specific behaviour]?			

The narrative should explain the choices made, addressing:

- Making maximum use of existing data – including a review of the quality and relevance of this
- Appropriate sampling – whether of people, sites, organisations or time periods – what type of sampling has been chosen and why this is appropriate for the type of generalization that will be undertaken.
- Appropriate data collection methods – why these methods have been chosen
- Appropriate data analysis methods – why these methods have been chosen

(ii) Checking the adequacy of the design in terms of evaluative questions

Many evaluations do not make explicit how evaluative questions will be answered – what the criteria will be (the domains of performance), what the standard will be (the level of performance that will be considered adequate or good), how different criteria will be weighted. A review of the design could check each of these in turn:

- Are there clear criteria for this evaluative question?
- Are there clear standards for judging the quality of performance on each criterion?
- Is there clarity about how to synthesize evidence across criteria? For example, is it better to have some improvement for everyone or big improvements for a few?
- Are the criteria, standards and approach to synthesis appropriate? What has been their source? Is further review of these needed? Who should be involved?

Ideally an evaluation design will be explicit about these, including the source of these criteria and standards. They might be set out in a table such as the following.

Table 1: Example table setting out the evaluative criteria, standards, synthesis process and sources

Evaluative aspect	Process for developing agreed standards, criteria and synthesis	Criteria	Standards	Synthesis/Weighting
Adequacy of resources for the program	Using national standards for the provision of services	Number of [services] per 100,000 people	[x] per 100,000 people	Average across all regions, weighted for population
Quality of services provided	National Service Standards	Financial accessibility	All people able to access services regardless of ability to pay	
Cleanliness	Food handling surfaces free from contamination			
Community consultation	Cultural appropriateness	People from all ethnic backgrounds feel welcome in the service		

(iii) Checking the adequacy of the design in terms of causal questions

Many evaluations do not make clear how causal questions will be answered. There are many designs and methods that can be used, but they involve one or more of these strategies:

(a) Compare results to an estimate of what would have happened if the program had not occurred (this is known as a counterfactual).

This might involve creating a control group (where people or sites are randomly assigned to either participate or not) or a comparison group (where those who participate are compared to others who are matched in various ways). Techniques include:

- *randomised controlled trials (RCTS)* – a control group is compared to one or more treatment groups
- *matched comparisons* - participants are each matched with a non-participant on variables that are thought to be relevant. It can be difficult to adequately match on all relevant criteria
- *propensity score matching* – creates a comparison group based on an analysis of the factors that influenced people’s propensity to participate in the program

- *regression discontinuity* - compares the outcomes of individuals just below the cut-off point with those just above the cut-off point.

(b) Check for consistency of the evidence with the theory of how the intervention would contribute to the observed results

This can involve checking that intermediate outcomes have been achieved, using process tracing to check each causal link in the theory of change, identifying and following up anomalies that don't fit the pattern, and asking participants to describe how the changes came about.. Techniques include:

- *contribution analysis* – sets out the theory of change that is understood to produce the observed outcomes and impacts and then searches iteratively for evidence that will either support or challenge it.
- *key informant attribution* – asks participants and other informed people about what they believe caused the impacts and gathers information about the details of the causal processes
- *qualitative comparative analysis* - compares different cases to identify the different combinations of factors that produce certain outcomes
- *process tracing* - a case-based approach to causal inference which focuses on the use of clues within a case (causal-process observations, CPOs) to adjudicate between alternative possible explanations. It involves checking each step in the causal chain to see if the evidence supports, fails to support or rules out the theory that the program or project produced the observed impacts
- *qualitative impact assessment protocol* – combines key informant attribution, process tracing and contribution analysis, using interviews undertaken in a way to reduce biased narratives

(c) Identify and rule out alternative explanations

This can involve a process to identify possible alternative explanations (perhaps involving interviews with program sceptics and critics and drawing on previous research and evaluation, as well as interviews with participants) and then searching for evidence that can rule them out.

While technical expertise is needed to choose the appropriate option for answering causal questions, as manager you should be able to check there is an explicit approach being used, and seek technical review of its appropriateness.

Causal relationship (between one variable and another – one step in the causal chain)

What strategies and methods/designs are being used for causal inference

eg Participation in program and improved health and wellbeing

Counterfactual – matched comparison groups of participants and non-participants

eg Increased skills and changed behavior

Consistency of evidence and ruling out alternatives –
process tracing and key informant attribution

(iv) Check that the design and process answers the action components of KEQs

Answers to action questions are often made in the form of recommendations. These don't necessarily flow straight from the findings. They often need an additional step of identifying possible actions and selecting the most appropriate, given the particular values and the availability of resources.

As manager you should check there is an explicit process for developing and reviewing recommendations, with appropriate levels of input from key stakeholders.