

CONTENT

STEP 5.2: OUTCOME & IMPACT EVALUATION

Evaluation terminology	2
a) Randomized Control Trials	3
b) Quasi-experimental design	3
When to identify comparison groups	3
Requirements and steps to implementing the quasi-experimental design	4
Steps for quasi-experimental data collection	4
Using vulnerability scores based on selection criteria to measure the similarity between communities	5
c) Value for money studies: cost-efficiency, effectiveness, and benefit studies	5
d) Success case method	6
Experience from Ethiopia	6
e) Qualitative Impact Protocol (QulP)	6
f) Participatory Impact Assessment	6
g) People-first Impact Method (P-FIM)	7
h) INDABA	7

Step 5.2: Outcome & Impact Evaluation

Outcome and impact evaluations seek to determine whether early actions helped people to prepare for or cope with hazard impacts during or immediately after a hazard event (see the box on evaluation terminology below for definitions and clarifications on evaluation terminology and anticipatory action).

Outcome and impact evaluations seek to determine whether early actions helped people to prepare for or cope with hazard impacts during or immediately after a hazard event (see **Box 1: Evaluation terminology** below for definitions and clarifications on evaluation terminology and anticipatory action). Although there are too many approaches to outcome and impact assessment to list here, the subsections below introduce methodologies that have been used or may hold promise for future evaluations of anticipatory action. Where applicable, we provide lessons from National Societies who have attempted them as well as links to additional resources.

While your National Society may be able to execute some of these approaches using staff and volunteers, all will produce less biased results if implemented by an external team and will require—at the very least—support from your PMER department. Support may also be available from a Partner National Society, the Climate Centre, or another reference center. If you are working with or on an s/EAP, the s/EAP monitoring and evaluation budget cannot exceed 5% of the total budget. This implies that most of these methodologies will require additional funding. If your National Society would like to undertake a more complex evaluation, contact your regional [DREF focal point](#) to explore the use of the DREF coordination budget to support this work. If your National Society attempts a methodology not listed here, please [contact GRC](#) to have your experience added to the manual.



Evaluation terminology

Evaluators typically refer to three different categories of results. The definitions below use examples from anticipatory action and align to the definitions found in the [IFRC M&E guidance](#) (p. 77-82).

1) Outputs: Measuring outputs is generally easy: keep track of trainings held, households reached, or kits delivered. This should be available through standard record keeping, such as distribution and participant lists. For anticipatory action, output measurement should also include *when* these items were provided or delivered. Because they are easy to measure, it is often referred to (somewhat pejoratively) as “bean counting:” they tell you who was reached, with what, and when but nothing about what was done with the services, information, or materials provided.

2) Outcomes generally refer to specific changes or immediate benefits resulting from program outputs and activities. In the case of anticipatory action, outcomes often manifest as what people do that they would not have otherwise because of the anticipatory support provided. Measuring outcomes of a shelter program for example, would entail determining whether households made the effort to strengthen their houses after receiving a shelter strengthening kit and participating in training on how to use it.

Tips for tracking outcomes:

If your National Society plans to conduct an outcome or impact evaluation, it will be important to identify groups of households who received *exactly the same* “package” of aid so that you can compare them to other groups who received different packages (or nothing at all). National Societies that have households sign for individual items rather than complete aid packages often have difficulties identifying exactly who received what. This greatly complicates sampling or selecting households for surveys, interviews, focus groups, or other data gathering activities. Where possible, develop distribution tracking systems that capture everything a household receives or allow you to easily cross-reference multiple lists to develop a master database.

3) Impacts are actual changes in livelihoods, well-being or other stated goals because of anticipatory actions. Impacts can only be measured after an intervention is complete. Given the relatively short duration of anticipatory action interventions and of the window of intended benefits, this may be as little as two weeks after an anticipatory action activation. Whereas outcomes measure whether people wash their hands more after a WASH training or how people spend anticipatory cash, impact measurement would determine whether they are less likely to fall ill because they are now washing their hands or whether they have better food security scores because of purchases made with the cash.

Outcomes vs impacts in anticipatory action: as in many discussions of terminology, exact usage and definitions often differ. Impacts as defined by the OECD/DAC are “positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended” ([OECD Glossary](#) p. 24). This definition reflects that the impacts of longer-term programs often cannot be measured in the short-term. For this reason, some practitioners think of outcomes and short-term results and impacts as long-term results. However, most anticipatory action initiatives are not seeking to effect long-term change (yet). Using the short- vs long-term definitions, outcomes would include both actions or changes facilitated by anticipatory support and any short-term impacts resulting from those changes. Nevertheless, the distinction between actions or other intermediate goals (e.g. using aid to prepare for a hazard based on a forecast by planting or harvesting crops, reinforcing houses) and the intended impact (e.g. a reduction in food insecurity, less damage to houses) is still important to understanding

whether anticipatory action initiatives are meeting their goals. For this reason, this guide uses outcomes to describe changes in knowledge, resources, or action resulting from anticipatory assistance and impact to indicate a reduction in negative (or an increase) in positive results (not based on measurement of longer-term results or sustained impact).

a) Randomized Control Trials

Randomized control trials (RCTs) are a highly robust method widely used in academia and in practice for impact evaluation. In an RCT, vulnerable households will be randomly assigned to receive one of the intervention packages to be compared. Comparison groups allow evaluators to establish what academics and evaluators call a **counterfactual**, or an understanding of what would have happened with a different intervention or the lack of intervention. Relevant counterfactuals for anticipatory action include different packages of anticipatory aid, the same package delivered at different times (including different anticipatory timings and/or response), or (ambitiously) Outcomes and impacts are then compared across the groups, enabling evaluators to attribute differences to the intervention (assuming the recipient and non-recipient households have similar socio-economic characteristics and are similarly at-risk/affected by the hazard). Note that either the aid package or the timing must be the same for two groups to be compared, otherwise it is impossible to determine whether the package or the timing were responsible for any differences in outcomes and impacts.

While RCTs produce the most rigorous evidence, they come with tradeoffs. Designing and implementing RCTs for anticipatory action requires considerable technical expertise and consequent financial resources. If your National Society is interested in exploring an RCT, it should seek support from the Federation, external researchers, or consultants with expertise in applying RCTs in humanitarian or development settings. [J-PAL](#), for example, provides resources on RCT and has considerable expertise in conducting them in humanitarian settings.

Ethical considerations may also play a role in decision-making. Some practitioners express ethical concerns about withholding aid from at-risk households. However, as organizations rarely have sufficient resources to provide anticipatory assistance to all households that might be affected by impending events, randomizing who among the most vulnerable receives assistance need not reduce the number of households reached overall. Using this method to compare anticipatory aid to traditional, post-response aid is another way to reduce such concerns. Nevertheless, randomization is significantly different from the community-based approach to recipient selection National Societies usually take when working with s/EAPs or sometimes also community-based approaches, thereby requiring considerable adjustment of normal protocols.

No National Society has attempted an RCT, other organizations engaging with anticipatory action have. A few other examples are available in the [toolbox](#) though.

b) Quasi-experimental design

Quasi-experimental studies are another way to use comparison groups to determine the outcomes and impacts of anticipatory action. As with RCTs, this methodology quantitatively compares recipients of anticipatory action to households who receive nothing, receive response assistance, or who received different packages of anticipatory aid. Rather than randomly assigning households to intervention groups, evaluators identify comparison communities during or after early action implementation (see Box 2) and randomly select samples from relevant groups or communities for comparison. Demographic, socio-economic, and vulnerability characteristics are then used to control for differences between the two groups. Box 3 outlines the key requirements of this methodology as well as the steps required for implementation. Box 4 summarizes the sequence of data collection. Additional guidance on this methodology can also be found in WFP's Guidance on [Monitoring and evaluation of anticipatory actions for fast and slow-onset hazards](#).



When to identify comparison groups

When to identify comparison groups will depend on the time you have during the implementation phase of early actions (i.e. whether your hazard is fast or slow onset) as well as the groups you are comparing to. For example, for drought activations WFP regularly identifies households that receive aid at different times and will collect baselines information for both groups. This enables them to compare outcomes over time. This is particularly valuable and feasible when National Societies know they will be providing response aid, as they can approach communities for baseline information without worrying about raising expectations, as both anticipatory action and response groups will eventually get assistance, only at different times.

Demographic, socio-economic, and vulnerability characteristics are then used to control for differences between the two groups.

The two infoboxes below outline the key requirements of this methodology and the steps required for implementation as well as the sequence of data collection. Additional guidance on this methodology can also be found in WFP's Guidance on Monitoring and evaluation of anticipatory actions for fast and slow-onset hazards that is included in the [toolbox](#).



Requirements and steps to implementing the quasi-experimental design

Before deciding to proceed with a quasi-experimental quantitative impact evaluation, ensure that your National Society will be able to do or provide the following:

1. Provide the number of people targeted in each community, including **what** they received and **when** they received it. The template available in the [toolbox](#) may help to compile this information.
2. A complete list of all the recipient households (i.e. the sampling frame). A corresponding map and/or contact information, while not required, will help you to more easily locate and survey the households and to sample in a way that controls for geographic location, if necessary.
3. Identify comparison communities that have similar characteristics to the recipient communities: the same livelihoods/socio-economic characteristics, the same exposure and vulnerability to the hazard and that have experienced the same intensity of impacts as the recipient communities. Guidance on selecting comparison communities is available in the [toolbox](#).
4. A complete list or map of households living in comparison households (those not receiving aid, receiving a different package, or receiving the aid at a different time). This will serve as your sampling frame for the comparison group(s), allowing you to randomly select comparison households. Again, having a list and a map and/or contact details will be helpful in locating the randomly selected households.

A note on comparing to non-recipient communities or households: when this methodology was attempted in Honduras and Ecuador, the National Societies found it difficult to obtain a complete list of households in non-recipient communities. No one (including community leaders) had access to complete overviews of community information. Both community leaders and the Red Cross were understandably reluctant to contact or gather people to develop comprehensive lists, as this raises expectations of aid. Unfortunately, the unavailability of lists or maps prevented the study team from getting a random sample, resulting in questionable data. For this reason, National Societies may find it easier to compare outcomes and impacts of aid received at different times.

5. The ability to collect data while outcomes and impacts can still be measured. The timing of data collection should be based on when you expect the impact of your intervention to materialize and how long they can be expected to last. Additional guidance can be found in the WFP guidance in the [toolbox](#).



Steps for quasi-experimental data collection

The following are the high-level steps for collecting quasi-experimental survey data. They are meant only to help you understand the process and decide if it is feasible for your National Society. If you are considering this methodology, you should also consult the guidance in the [toolbox](#) and seek external support as required.

- Decide who will collect data (National Society volunteers or external parties)
- Decide who will analyze data (National Society volunteers or external parties)
- Develop survey (examples from 4As and GRC library)
- Obtain lists and/or maps of all recipient and comparison/non-recipient communities (sampling frame).
- Use sample size calculators to determine how many households must be surveyed. Most multi-community interventions will require cluster sampling. We recommend you consult with a statistician to ensure you sample correctly and get a large enough sample size.
- Draw a random sample from the intervention and comparison groups.
- Create budget (sample budget from other contexts). **Note:** it is important to budget enough time and resources to reach ALL the households that appear on these random lists. This should be done by visiting them in their homes rather calling them to a single meeting point. Calling them to one place increases the likelihood of bias. Failing to speak to even one household introduces possible bias, making it difficult to generalize the results.
- Train the data collectors.
- Test the survey in the field and update the survey. **Note:** It is important to test the tool and make sure it is working correctly after each and every change to the tool.
- Collect data from all recipients **before** starting data collection for the non-intervention and/or response group. This allows you to calculate a vulnerability score (see **Box 5**) and ensure that respondents from non-beneficiary households are comparable.

- After data from the recipients is complete, calculate the vulnerability scores and identify a range of scores that participants from comparison communities should fit it.
- Collect data from non-beneficiary households using the vulnerability score, again without omitting any households.
- When possible, qualitative data (from key informant interviews and/or focus group discussions) collection should be conducted at the same time to help interpret the quantitative results.
- Analyze the data & write the report

Two National Societies have successfully produced three quasi-experimental impact studies of anticipatory action: Mongolia and Bangladesh (see [toolbox](#)). Other National Societies have attempted this methodology, but encountered challenges, particularly in obtaining random samples.



Using vulnerability scores based on selection criteria to measure the similarity between communities

In quasi-experimental designs, it is necessary to ensure that the households you are comparing have similar socio-economic and demographic characteristics and that they have experienced the hazard in the same way. For example, you should not compare households with large livestock holdings to those who do not have any livestock or communities/households that can access irrigation during a drought to those who must walk several kilometres to get water. Often, the selection criteria for providing anticipatory assistance are an excellent starting point for ensuring that the populations are similar, as they point toward the critical factors contributing to vulnerability.

Based on the selection criteria and indicators of relative poverty or income appropriate for your context (such as livelihood and incomes sources or types of housing construction), you and your MEAL team will design survey questions relevant to each factor contributing to household vulnerability and assign relative scores to all possible answers. In this way, each household will be assigned a relative vulnerability score based on their circumstances. Sample vulnerability scores can be found in the [toolbox](#).

Having and applying clear selection criteria when selecting recipient households will make it easier to establish questions and scores for your quasi-experimental impact evaluation.

c) Value for money studies: cost-efficiency, effectiveness, and benefit studies

Given an increasingly tight funding climate and interest from National Society partners, your National Society may wish to demonstrate the value for money of specific programmes or actions. Early attempts to measure the return on investment were not particularly transparent; however, emerging guidance (see toolkit) on how to conduct transparent rigorous value for money analyses of anticipatory action provides a good starting point for National Societies looking to conduct such analyses. Examples of costing studies can be found in the [toolbox](#).

Cost-efficiency, cost-effectiveness, and cost-benefit analyses are three different but related ways to demonstrate the financial benefits of programs. Each of the three methods below will require detailed analysis of programme budgets to determine the staff, overhead, and procurement costs that went into anticipatory action (and any alternative programmes to which it is being compared). They thus lend themselves more for use in the context of formalized programmes or protocols.

- **Cost-efficiency analysis** measures how much it costs to deliver a unit of output (e.g. cost per water purification kit delivered). It is useful for comparing the costs of delivering the same aid package at different times (e.g. anticipatory action vs response). Web-based programmes such as [Dioptra](#), have developed standardized methodologies and tools to help aid organizations measure and compare the cost effectiveness of different interventions.
- **Cost-effectiveness analysis** measures how much it costs to generate a relevant outcome (e.g. cost per unit of improved food consumption score). This approach is useful when trying to compare outcomes or impacts that are difficult to monetize, such as loss of education or increases in protection. It can also be used to compare the costs associated with achieving the same outcome using different interventions.
- **Cost-benefit analysis** monetizes the values of outcomes or impacts and calculates whether those monetized benefits are greater than the costs. These kinds of analyses can produce a benefit-cost ratio, a net present value, and return on

investment calculations. Such studies can be used to justify investments in anticipatory action

Note/Tip: To perform rigorous cost-effectiveness or cost-benefit analysis, you must first conduct a study to measure the outcomes and/or impacts. These studies are therefore only as rigorous as the outcome or impact studies supporting them.

A word of caution: In practice, the organization or consultant conducting the cost-benefit study will decide which costs to include and exclude (e.g. the value of setting up the programme or only the implementation of early actions itself) and how to monetize benefits. Because the final cost-benefit or return on investment numbers may be quite sensitive to these decisions or assumptions, being transparent about how and why particular numbers or assumptions were chosen will help to lend credibility to the results.

Additional resources on value for money in disaster risk finance and how to conduct cost-efficiency, -effectiveness, or -benefit analyses can be found in the [toolbox](#).

d) Success case method

The success case method has been identified as a promising mixed-methods methodology for National Societies desiring some quantitative measurement of outcomes but without the capacity to implement an RCT or quasi-experimental survey. The approach begins with a survey that includes demographic and vulnerability questions as well as critical outcome and impact indicators. The survey data is then analyzed to identify households who benefitted more (“success cases”) and less (“non-success cases”) from the interventions. The research team then follows up with several households to conduct in-depth interviews aimed at understanding why those households had comparatively better or worse outcomes. If resources allow, it can be expanded to include data collection from non-recipient or other comparison groups. Guidance on how to adapt the success case methodology to anticipatory action can be found in the [toolbox](#).

Note/Tip: If you are considering the success case method, your PDM survey is an excellent opportunity to collect survey data and identify households who benefitted more (“success cases”) and less (“non-success cases”) from your interventions. By adding questions on key impact indicators and requesting contact information and permission to follow-up for in-depth interviews, your PDM could fulfill the survey component without requiring a separate survey.



Experience from Ethiopia

Ethiopia Red Cross Society was the first to test this methodology in 2025. Tools used by Ethiopian Red Cross can be found in the [toolbox](#).

e) Qualitative Impact Protocol (QuIP)

The Qualitative Impact Protocol is a methodology designed to attribute change (i.e. outcomes and impacts) to an intervention through narratives, stories, and in-depth interviews with intervention recipients. Although it has been successfully used by humanitarian actors, it requires highly trained, independent data collectors, meaning that it may be challenging for National Societies with scarce resources. This method has not yet been used to evaluate anticipatory action programmes. More information on this methodology can be found in the [toolbox](#).

f) Participatory Impact Assessment

Participatory impact assessment (PIA) is based upon participatory rural appraisal (PRA) which is similar in approach and philosophy to Red Cross Red Crescent vulnerability and capacity assessments. It has three main objectives:

- to identify the factors leading to change in people’s lives;
- to determine which of these factors can be linked to a specific intervention;
- to determine the importance of each factor.

PIA allows communities to participate in the definition of indicators rather than relying on pre-determined or prescribed targets. Rather than attempting to compare results to non-participants in a programme, it asks participants to describe change over time. Many activities begin with a ranking or scoring activity, followed by an open-ended discussion to understand the results of the ranking. While the activities are often considered qualitative, when repeated in a standardized way, the ranking and scoring activities can yield some numerical data and quantitative insights or “participatory numbers.” This methodology has been used outside of

anticipatory action to evaluate livestock interventions, particularly in pastoralist settings. A guide to participatory impact assessment can be found in the [toolbox](#).

g) People-first Impact Method (P-FIM)

The people-first impact method (P-FIM) is a community engagement tool that respects the right and ability of communities to take the lead in the design, implementation, and evaluation of aid programs. Ideally, it would be employed from the start (for example at feasibility study and EAP design stage), to ensure that anticipatory action initiatives in general, and specific actions in particular, address genuine community needs and integrate community abilities and solutions in an added-value cost-effective and sustainable approach. P-FIM builds trust, two-way engagement, and transparency between agencies and communities. It identifies the most critical areas for community-led action without introducing organizational bias or pre-conceived and (possibly) untested assumptions. Through two-way dialogue facilitated by trained local teams who live and work in the context, P-FIM brings to light what people believe are the most critical issues and forces impacting their lives, positive and negative. It is valuable in identifying appropriate interventions and adds value and relevance to existing agency tools by allowing communities to lead in the selection and development of MEAL plans and tools, such as theories of change or post-distribution surveys.

As a community engagement approach, P-FIM fits naturally with MEAL: engaging the community allows communities to shape MEAL components, ensuring they are easily understood by communities and agency frontline staff and volunteers. P-FIM challenges agencies to develop simple easily understood tools. Like PIA, it helps agency staff and volunteers understand what communities and organizations feel should be monitored and evaluated and how. In the P-FIM approach, the evaluation questions used in focus group, key informant interviews, and other discussions are defined by communities themselves with the National Society. Similarly, P-FIM engages communities in selecting and designing other (perhaps more academic) methods to be used for evaluation.

Welthungerhilfe uses the method to inform its anticipatory action programs. The [toolbox](#) includes a guide explaining the full P-FIM approach and an overview of how P-FIM two-way engagement can be applied to MEAL at all stages of a programme. Within those documents, you will find contact information to request a two-day, practical P-FIM training for your team.

h) INDABA

INDABA is an innovative participatory video process that allows communities to design, collect, edit and develop their own stories through a mobile application that can be used onsite and in remote contexts. IFRC's Strategic Planning Department in coordination with National Society counterparts, developed and is supporting this qualitative approach in PMER for long term and emergency contexts. It has been rolled out in Colombia, Egypt, Eswatini, Gambia, Honduras, Indonesia, Kenya and Namibia. Examples can be found on IFRC's [PMER \(Planning Monitoring Evaluation and Reporting\)](#) YouTube page.

This qualitative approach relies strongly on community engagement in the process to allow for feedback not only at the community level, but also at the level of disaggregated groups (e.g. the elderly, women, men, and youth). The process is adaptable to context and uses different facilitation methods (photos, story boarding cards, sectorial cards with questions, tags, dice etc.) to engage participants, put them into the context of the project or programme and allow them to share feedback that can inform improvement. The online process has been simplified to allow community members to easily design, tag, edit and create their own videos.

Although not yet explicitly applied to anticipatory action, the INDABA method that could be relevant to anticipatory action for the following reasons:

- It is a community-based approach. Communities are the first to detect early warning signs of crises;
- The feedback collected through video stories can provide the IFRC network, partners and donors with real time feedback on anticipatory action rooted in local knowledge.
- Story telling is a strong tradition in many cultures. The production and sharing of video stories could allow people to describe their experience of anticipatory action in their own ways. It could also help promote peer-to-peer knowledge transfer in anticipatory action.