USAID
FROM THE AMERICAN PEOPLE

EVIDENCE IN ACTION

USING AND GENERATING EVIDENCE ABOUT EFFECTIVENESS IN BIODIVERSITY PROGRAMMING

Unit 3: Generating Evidence

APRIL 2018

Prepared by: Natalie Dubois, Environmental Incentives; Andrés Gómez, ICF

Front Cover: Implementing activities provides an opportunity to generate evidence about the effects of different fishing practices on wild fish populations, such as the shallow water nets being used by women fishing in Tanzania. Photo credit: Matt Kieffer

Back Cover: A view of primary rainforest in Honitetu village, West Seram regency, Maluku province, Indonesia Photo credit: Ulet Ifansasti/CIFOR

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

Acknowledgments: The authors would like to acknowledge Diane Russell for her vision, guidance, and contribution to *Evidence in Action*. Elizabeth Daut provided helpful comments on a draft of this unit.

About Measuring Impact: This work is made possible by the generous support of the American people through the United States Agency for International Development (USAID) under the terms of its requisition number REQ-EGAT-12-000014 (Measuring Impact) implemented by Environmental Incentives, LLC; Foundations of Success; and ICF International. Measuring Impact has been issued under contract number AID-OAA-C-12-00078 and supports the same program objectives as described in RFP number SOL-OAA-000050. Measuring Impact is funded and managed by the USAID Bureau for Economic Growth, Education, and Environment Office of Forestry and Biodiversity.

TABLE OF CONTENTS

LIST OF BOXES

LIST OF EXAMPLES

LIST OF TABLES

ACRONYMS

USAID	United States Agency for International Development
MEL	Monitoring, Learning, and Evaluation

1. OVERVIEW

Healthy rivers, forests, and oceans are essential to development, as they support and sustain livelihoods and human well-being. Conservation protects the biological resources that people depend on and that are a critical component of good development outcomes. To this end, the United States Agency for International Development (USAID) has made significant investments in mitigating threats to biodiversity in key ecosystems and landscapes.

Faced with finite resources and great demand, it makes sense to ask tough questions about the effectiveness of biodiversity programs. It is not only important to know if a program achieved its expected outcomes; it is also important to understand how and why a program achieves success. Using and generating evidence about what works, what doesn't, and in which contexts can help teams make better programming decisions. *Evidence in Action* helps mission staff and implementing partners use and generate evidence about the effectiveness of biodiversity programs. The resource is presented in four units that can be used alone or as a series. A glossary defining key terms is included with each unit.

- *Unit 1: Understanding an Evidence-Based Approach* provides an introduction to evidence and evidence-based approaches to biodiversity programming in the context of the USAID Program Cycle.

- *Unit 2: Using Evidence* focuses on the critical review and use of evidence to increase the effectiveness of biodiversity programs.

- **This third unit: *Generating Evidence* identifies Program Cycle processes that teams can use to generate credible evidence about the effectiveness of biodiversity programs.**

- *Unit 4: Building the Evidence Base* highlights ways in which evidence can be shared and applied to strengthen biodiversity programs across USAID.

2. INTRODUCTION

In a perfect world, USAID staff and implementing partners would have all the evidence (see Box 1on page 7) needed to make decisions: they would know that the diagnosis of the problem was accurate, that the strategic approach was effective, and that the assumptions in the theory of change would hold in the program context. But frequently, teams lack the information they need to make informed decisions. This unit of *Evidence in Action* explores actions that teams can take to generate their own evidence to address these information needs.

Without evidence about what works and what doesn't work, teams are susceptible to missing opportunities to replicate successes by continuing to invest in programs with a low track record of success. To address information gaps, teams can consider generating evidence that will increase their understanding of three components of program success:[1]

- The accuracy of the **problem analysis** *(Is the diagnosis of the problem and assessment of the context correct?)*
- The validity of the assumptions in the **theory of change** *(Is the understanding of how change happens correct?)*
- Appropriate **program implementation**? *(Are the actions being used to implement the strategic approach appropriate?)*

Once a team is aware of these information needs and the points at which they are likely to be identified in the Program Cycle, they can consider generating evidence to address them.

Unit 3: Generating Evidence covers topics related to generating evidence about the effectiveness of biodiversity programs. The unit is organized around three topics:

1. Setting priorities for generating evidence about program effectiveness
2. Selecting an approach for generating evidence about program effectiveness
3. Generating and interpreting evidence about program effectiveness

BOX 1: WHAT IS EVIDENCE?

Automated Directives System (ADS) Chapter 201 defines evidence as the "[b]ody of facts or information that serve as the basis for programmatic and strategic decision-making in the Program Cycle. ... [Evidence] can be sourced from within USAID or externally and should result from systematic and analytic methodologies or from observations that are shared and analyzed" (page 145).

The term "evidence" refers to both (1) individual findings or pieces of information used to help make a decision or support a conclusion; and (2) the body of findings or information providing support for (or countering) a belief or claim.

Evidence can be generated through primary research, literature reviews, case studies, assessments, evaluations, and performance monitoring. Evidence for program effectiveness comes from real-world observations and documentation of program outcomes. Observations are not considered evidence unless they are used to investigate whether a belief or claim is true.

After completing this unit, teams will be able to:

- Prioritize information gaps they are likely to encounter in the design and implementation of strategic approaches
- Use existing mechanisms to generate evidence in support of more effective programs
- Describe basic data collection and analytic designs that are appropriate for generating evidence about the effectiveness of programs

Testing the validity of assumptions about the drivers of threats posed to gorillas by human populations is an important part of understanding conservation outcomes in USAID programs in Uganda. Photo credit: Douglas Sheil/ CIFOR

3. SETTING PRIORITIES FOR GENERATING EVIDENCE

How does a team decide which topics should be a high priority for generating evidence? *Unit 2: Using Evidence* focused on using the existing evidence base to address important information needs. By examining the evidence base at the beginning of a project, program managers and implementing partners try to reduce some of the uncertainty associated with program assumptions, thus increasing the likelihood of program success. However, teams will often discover that the existing evidence base does not fully meet their information needs. This unit, *Generating Evidence*, focuses on how teams can fill those gaps to increase their understanding of program effectiveness.

Three broad priorities for generating evidence about the effectiveness of biodiversity programs are strengthening:

1. **The problem analysis** – Testing the validity of assumptions in the problem analysis where there is low confidence that the correct drivers have been identified
2. **The theory of change** – Testing the validity of key assumptions in the theory of change that have a limited evidence base
3. **Implementation** – Assessing the appropriateness of actions intended to achieve desired results in the program context

The approach for identifying key assumptions that was presented in *Unit 2: Using Evidence* (see Section 3 in Unit 2) can also be used to identify priorities for generating evidence about assumptions that have a weak or limited evidence base in the program context (see Box 2 on page 9).

An essential component of the learning sections in MEL Plans should focus on: "Identifying knowledge gaps during strategy development or project design and implementing plans to address them through evaluations, use of monitoring data, or other means" (ADS Chapter 201, page 133).

BOX 2: ASSUMPTIONS THAT ARE PRIORITIES FOR GENERATING EVIDENCE

Certain assumptions are particularly important because they are likely to jeopardize program success if invalid. Testing the validity of these assumptions should be a high priority for generating evidence when uncertainty about their validity has not been resolved during program design. These include:

1. Assumptions in the problem analysis that identify the driver that a strategic approach is designed to directly influence
2. Assumptions in the problem analysis that identify the immediate cause of the threat
3. Assumptions with doubtful causality in the theory of change
4. Assumptions about the effectiveness of actions intended to influence key drivers in the program context

Reducing community reliance on protected forest resources may depend on the success of propagation efforts like those shown for these Allanblackia *seedlings at a nursery in Ghana. Photo credit: Cyril Kattah*

PRACTICAL TIPS FOR PROGRAM MANAGERS

Program managers may wish to revisit the key assumptions identified in Box 2 on page 9 to identify gaps where the evidence base is weak or uncertain in their context. By generating evidence to assess the validity of key program assumptions (see Table 1), teams can increase their understanding of what approaches are likely to work, and the conditions that need to be in place for them to work, in particular contexts.

As teams review their information needs, they should be careful to distinguish where performance measures – which provide evidence that certain conditions exist or that certain results have or have not been achieved – are appropriate and where additional data or different data designs are needed to test the validity of assumptions underlying a program's effectiveness (see "Using Monitoring Information" in Section 5).

Table 1: Examples of information needs that focus on the validity of a program assumption and potential priorities for generating evidence.

Information need	Example
In the problem analysis, uncertainty about the cause of the problem can undermine program success when it results in the team addressing an incorrect driver.	A team is planning to invest heavily in a strategic approach to combat wildlife crime but is uncertain whether existing social norms are the main driver of consumption of illegal wildlife products.
In the problem analysis, uncertainty about the contribution of identified threats and drivers to the status of the biodiversity focal interest can undermine program success when it results in the team focusing on minor contributors to the problem.	A team is confident that the strategic approach will lead to fewer individuals consuming illegal wildlife products, but is uncertain whether local consumption is a significant contributor to sales of illegal wildlife products.
In the theory of change, uncertainty about how change occurs can undermine program success when it results in the team pursuing ineffective solutions.	A team is using behavior change methodologies to reduce consumer demand for illegal wildlife products but is uncertain whether providing information on the negative impacts of using illegal wildlife products will affect consumption patterns in local communities.
In implementation plans, uncertainty in the effectiveness of proposed actions can undermine program success when it results in the team using actions that are unlikely to achieve change in key drivers.	A team has designed an awareness campaign that relies on program staff to deliver key messages, but is unsure whether stakeholders will be receptive to messages perceived as coming from outside the community.

4. SELECTING AN APPROACH FOR GENERATING EVIDENCE

There are three general approaches within the Program Cycle that program managers can consider using to generate evidence about program effectiveness. Regardless of the approach, teams will need to articulate questions that elicit evidence that addresses the information needs that they have identified as priorities.

1. **Commissioning research through existing mechanisms or partnerships.** Commissioned research can be a particularly relevant approach for testing assumptions in the problem analysis.

2. **Designing evaluation questions to strengthen understanding of the theory of change and its implementation in the local context.**

3. **Collecting relevant data as part of monitoring during implementation of activities**. Monitoring data are particularly helpful for testing assumptions in the theory of change and the appropriateness of actions taken to implement the strategic approach.

COMMISSIONING RESEARCH

In some cases, the existing evidence base is insufficient, uninformative for the local context, or nonexistent. If this lack of evidence implies that there is significant uncertainty about the accuracy of the problem analysis and/or the strategic approaches being considered, then teams should consider funding specific research activities to address it.

Commissioning literature reviews or independent studies to address questions about assumptions in the problem analysis can be particularly valuable; that is, when a team recognizes that there is uncertainty in the identification of the drivers in the problem analysis. Having an accurate understanding of the local drivers affecting the status of biodiversity focal interests increases the likelihood that appropriate strategic approaches will be selected. For example, a team may consider commissioning a political economy analysis [2] to answer questions about who is engaging in specific behaviors and why, in order to better understand the components that can be influenced to achieve conservation outcomes.[2] This approach is most informative when timed to inform activity design.

Commissioning literature reviews or independent studies can also be useful when there is general uncertainty about the effectiveness of a strategic approach or the conditions under which it is most likely to work. This may be especially relevant for assessing the validity of the assumptions in the theory of change at the project level, before the effectiveness of particular strategic approaches are being considered within their specific activity contexts.

When considering commissioning research, teams should focus on a limited set of questions. It may be costly and time-consuming to manage a portfolio of several simultaneous research activities. The selected questions should be feasible to research with the time and resources available. Important considerations include the timeframe in which the decision must be made and the budget available for research.

Teams should focus their efforts on questions that can be expected to lead to actionable information. Questions that can be reasonably answered by "it depends" or "sometimes" and those that will not directly affect specific decisions are unlikely to be good candidates for commissioned research. Required assessments (such as the Gender Analysis, the Biodiversity and Tropical Forestry (Foreign Assistance Act 118/119) Assessments,[3] and Climate Risk Management) can help teams align commissioned research with known information gaps.

USAID commissioned a comprehensive literature review by researchers at the American Museum of Natural History (Sterling et al. 2016) that looked at the effectiveness of different methods of stakeholder engagement. Photo credit: Ulet Ifansasti/CIFOR

EVALUATION DESIGN

Evaluation at USAID generates evidence for accountability and for learning to improve development outcomes. Accountability measures how well program activities have met expected objectives. An evaluation draws conclusions about the quality, merit, or worth of the program at achieving its objectives, with a goal of attributing results to the program to the extent possible (see Box 3 on page 13). Learning aims to test the fundamental assumptions that underlie strategic approaches and program design. Learning is not a judgment of what worked and did not work, instead it is an empirical process to build understanding of what is likely to work and why.

Evaluations are particularly valuable for questions about the appropriateness of actions. Evaluations can be used to assess the relationship between program actions and results in a specific context. The evaluation design determines the extent to which causal relationships can be established beyond simply tracking results over time.

Evaluations can also be useful for questions about assumptions in the theory of change for a particular strategic approach. The purpose of testing the assumptions in a theory of change is to examine whether a team's understanding of the change process is correct. Teams should consider framing

BOX 3: TYPES OF EVALUATIONS AT USAID

According to the USAID Evaluation Policy, "[e]valuation is the systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments, to improve effectiveness, and/or inform decisions about current and future programming" (page 2). USAID uses two types of evaluations for its programs:

Impact evaluations are most often associated with testing causal relationships. They are designed to measure the change in a development outcome that is attributable to a defined intervention (USAID 2013), which may be a strategic approach or, more commonly, an activity comprising multiple strategic approaches. An impact evaluation focuses on a particular causal relationship, in which the activity (or strategic approach) is the independent variable and one or more specific outcomes of interest are the dependent variable. An impact evaluation often employs a rigorous design involving a counterfactual to control for factors other than the activity that might account for the observed change.

Performance evaluations typically focus on descriptive and normative questions, but they can also address – and typically do address – cause-and-effect questions (USAID 2016a). For example, a performance evaluation might be used to test a causal relationship between results in the theory of change. Typically, performance evaluations do not involve the use of a counterfactual, although other types of comparison groups may be used.

evaluation questions that can test "forward causality" (i.e., what is likely to happen if result X is achieved) not just attribution (i.e., what caused result Y).

When considering using an evaluation to generate evidence about program effectiveness, teams should focus on questions for which it is feasible to obtain appropriate data with the time and resources available. The team must be reasonably certain that they will have enough time to observe changes, and that they have sufficient resources to evaluate conditions across a large enough sample to render credible conclusions.

Teams should also assess whether appropriate study designs are feasible. Is it possible to generate information that will rule out alternative explanations and eliminate bias? The team should consider the strengths and limitations of both qualitative and quantitative designs.

USING MONITORING INFORMATION

Within the Program Cycle, monitoring and evaluation processes provide an opportunity to gather and analyze various data on programs, including actions, inputs, outputs, and outcomes. These very same variables frequently show up in questions that address the effectiveness of strategic approaches and underlying assumptions. As a result, monitoring, evaluation, and learning processes can often be used to generate evidence about the assumptions underlying program effectiveness.

Teams should be explicit in characterizing their information needs and the decision(s) that the evidence being generated is intended to address. Indicators that track inputs, outputs, and outcomes describe what is happening or what has happened where the program has been implemented. Taken alone, indicators do not generate evidence about causes or relationships.

A team measures tree circumference as part of the USAID-supported Sustainable Wetlands Adaptation and Mitigation Program in the province of Central Kalimantan, Indonesia. Photo credit: Nanang Sujana/CIFOR

Missions that are interested in generating more rigorous evidence about effectiveness may consider asking implementing partners to develop monitoring protocols that serve dual purposes to: (1) generate data that can be summarized as a performance indicator and (2) address effectiveness questions. Additionally, information from indicators can be combined with qualitative data to generate evidence exploring why a result has or has not been achieved.

Monitoring, evaluation, and learning processes, such as evaluations, learning reviews, and pause and reflect practices are particularly valuable for framing questions about the validity of assumptions in the theory of change that persist through activity start-up and early implementation. They also help teams assess whether the necessary enabling conditions are present in the program context. These processes and practices are helpful for determining when results are not on track (i.e., assumptions are not holding in the program context), so that teams can make changes in response to new information.

When considering using monitoring, evaluation, and learning processes to generate evidence about effectiveness, teams should focus on a limited set of questions that address key uncertainties about the assumptions in the theory of change. These are questions that go beyond simply establishing whether results have or have not occurred but instead focus on explaining how and why the strategic approach is expected to achieve those results.

The team should also consider where the activity is in the Program Cycle (Are there opportunities to collect additional data or modify monitoring protocols? Or do existing data limit the questions that can be answered about the assumptions in the theory of change?). Questions identified at the middle or end of a program's implementation may not be fully answerable with data that have been collected to fill other information needs.

For example, monitoring data may indicate that the percentage of patrols operating as scheduled has increased in targeted protected areas and that the number of verified poaching incidents decreased across the same time period. However, if the team is questioning whether increased patrol effort is what causes observed changes in poaching rates, they may not have enough information to assess this relationship without additional data allowing them to compare differences in poaching incidents across varying levels of patrol effort.

5. GENERATING EVIDENCE ABOUT EFFECTIVENESS

Implementing partners and contractors often undertake the process of generating evidence through various Program Cycle processes. This division of roles and responsibilities means that USAID program managers and implementing partners need to work together to clearly communicate their information needs. Specifically, program managers should carefully articulate the questions that implementing partners or contractors are expected to address. Ensuring that the team's questions are clearly articulated and feasibly researchable will increase the likelihood that relevant and credible evidence will be generated.

Many questions about program effectiveness aim to test the validity of specific assumptions. These questions often can be framed to be answered through investigation or observation. In this context, a testable question is designed to generate evidence that supports or refutes specific assumptions about how the program works. In *Unit 2: Using Evidence*, these types of questions were identified as "foreground" questions as distinguished from "background" questions (see Box 4 on page 19).

A testable question is specific. It elicits specific data needs relevant to a claim and identifies what to measure and on what or whom.

A testable question lends itself to one or more falsifiable hypotheses. A hypothesis is a tentative answer that can be verified by investigation or methodological observation. A hypothesis is falsifiable if it is possible to describe an observation or argument that would prove it wrong. If a hypothesis is not falsifiable it has no predictive or explanatory value because it is consistent with all possible observations. Teams should not pursue efforts to generate evidence around non-falsifiable hypotheses.

PRACTICAL TIPS: ARTICULATING TESTABLE QUESTIONS

A claim is a statement that describes what the team would expect to observe if the assumptions they are making in their problem analysis, theory of change, and implementation plan are true. When there is uncertainty as to the validity of a claim, it is considered a hypothesis and can be used to direct evidence generation. In some cases, an assumption must be unpacked into one or more claims in order to articulate a testable question.

Teams may need to generate evidence to test assumptions about the economic and other benefits provided by enterprises such anchovy production shown in Maluku province, Indonesia. Photo credit: Ulet Ifansasti/CIFOR

Consider an assumption identified in the Cross-Mission Learning Agenda for Conservation Enterprises:

If the enterprise generates revenues and is sustainable, then stakeholders will realize benefits (primarily a marginal increase in income, but also additional non-cash benefits).

Here the team might identify the following claims:

- Descriptive – Stakeholders participating in enterprises that generate revenues receive income from the enterprise
- Causal – Income received from participating in a conservation enterprise increases stakeholders' household incomes

If the team is uncertain as to the validity of these claims they would consider them hypotheses. Then they would frame questions based on those hypotheses:

1. *Do stakeholders participating in revenue-generating conservation enterprises receive income from the enterprise?*
2. *Does the total household income among participants in conservation enterprises increase significantly over time (relative to non-participants)?*[4]

When articulating testable questions, the team should consider defining, identifying, or articulating the following[5] to increase the clarity of their framing:

What is the subject or population of interest? In the two questions above the subjects are implied in each question (participants in conservation enterprises and matched groups of participants and non-participants, respectively).

What is the outcome of interest? In the two questions above the outcomes are explicitly included in each question (the distribution of revenues from conservation enterprises and the total household income, respectively).

For causal claims, what factors explain or predict the outcome? In the second question above participation in conservation enterprises is the variable assumed to explain any observed increases in total household income.

For causal claims, what comparison groups or additional information could be used to rule out alternative explanations for the outcome? In the second question above there is an explicit mention of a comparison group that rules out factors that affect household income regardless of participation status as the cause of change in household income among participants. Another approach would be to survey heads of households participating in the conservation enterprise to identify the factors to which they attribute the increase in household income.

BOX 4: QUESTIONS THAT DON'T TEST ASSUMPTIONS

- Not all questions test assumptions. In *Unit 2: Using Evidence,* these types of questions were described as "background questions" (see Section 3 in Unit 2). Background questions are asked to elicit possible assumptions rather than test an assumption that has already been identified.

- Background questions are useful when teams do not yet have enough information to articulate their assumptions or are exploring why a particular outcome did or did not occur. However, background questions are not an efficient means of generating evidence about the validity of a particular assumption. Since they are more exploratory in nature, background questions require different data collection methods than foreground questions focused on testing claims.

PRODUCING RELEVANT DATA

Program managers often identify the questions about program effectiveness that will be addressed through the procurement of research activities, activity evaluations, or learning reviews initiated by implementing partners. The role of program managers in these situations is to review the appropriateness of the approach used to generate data and the research designs used to address the questions posed. The ultimate aim is to produce data that are credible, generalizable, and directly relevant to the team's needs (i.e., not just a collection of anecdotes). Many issues of quality and strength of evidence can be addressed with appropriate data collection and research design. Section 5 in *Unit 2: Using Evidence* provides a synopsis of several important issues affecting the quality and strength of evidence that program managers should keep in mind when reviewing approaches that have been proposed for generating data.

While there is no single approach to a given question, ensuring that important parameters have been considered can produce more robust evidence for a particular assumption. A detailed description of research methods in conservation science is beyond the scope of *Evidence in Action*, but consideration of several key concepts can help managers avoid common mistakes in data collection. For instance, much applied research in conservation fails to define the target population, define key terms, or define a baseline.

Clearly define the population and sample. A population is a well-defined collection of individuals or other entities with similar characteristics. If the entire population will not be sampled, the methods used to select subjects should also be described. The population is important because it denotes the group to which the conclusions from the data can be applied. When not sampling the entire population, the size and representativeness of the sample are important considerations; samples should be as large as is feasible given the resources available and representative of the whole population along as many attributes as it is feasible to define. When defining subsamples, researchers need to be transparent about the instruments (or sample frames) they used to include/exclude subjects.

Define the relevant elements of the question. Key elements implicit or explicit in the assumption should be defined in as much detail as possible. Questions often include terms such as "improve," "enhance," or "impact." Leaving such terms undefined can lead to answers that lack precision and could therefore be uninformative. If the team is unable to define key terms and propose feasible measures to assess them, they should consider rewording the question.

The wording of the question can affect the quality of the evidence produced to answer it. Teams should consider framing the questions to unequivocally point to evidence that will be actionable in their context. Too often in the field of conservation and development, important evidence gaps are framed in questions that begin with "To what extent." This framing is usually uninformative unless the team is specifically pursuing a question about degree or scope. Framing questions about causality, effectiveness, and correlation in this manner can limit the usefulness of the evidence produced.

Robust evidence for causality requires consideration of multiple factors (see "Interpreting Evidence" below). Asking "to what extent does X lead to Y" is unlikely to be a testable question. Instead, teams may ask questions about time order (e.g., does Y change after X changes) or correlation (e.g., are changes in X associated with changes in Y) and produce data in ways designed to help eliminate alternative explanations.

Identify appropriate baselines. Assessing change is fundamental to a theory of change and is required to determine the validity of many assumptions. For example, a theory of change often identifies a series of results that are causally linked, i.e., the assumption is that a change in one result is what causes the change in a subsequent result. In order to establish change, there must be some baseline against which to compare the measured outcome. Teams must consider the sample size and representativeness of the population used to establish the baseline as well as the appropriateness of the length of time over which data will be collected.

A baseline coming from only a few select observations may provide a very unreliable picture. For instance, fishing patterns might be seasonably variable. If the baseline data do not accurately represent fishing across the seasons, teams may be blind to, underestimate, or overestimate the impact of their strategic approaches.

EXAMPLE 1: FORMULATING A TESTABLE QUESTION ABOUT THE RELATIONSHIP BETWEEN COMMUNITY ADVOCACY AND LAND USE PLANS

An implementation team is developing their monitoring, evaluation, and learning (MEL) plan for an activity that uses community advocacy as a way to influence and make land-use decisions in sensitive wetland areas. The mission has invested heavily in this strategic approach and wants to ensure that this activity allows them to generate evidence about its effectiveness.

The team starts with the following question: did community advocacy for stronger wetland protections lead to more protective zoning in land use plans? They decide to include it as a learning question. The implementing partner included indicators tracking participation in community advocacy and the quality of land use plans each year of the activity in their MEL plan.

Collaborative land use planning in Mamberamo Raya Regency, Papua, Indonesia. Photo credit: Mokhamad Edliadi/CIFOR

The following statement summarizes progress towards the established benchmark[6] identified by the activity team during activity start-up. It is based on indicator data collected during the period of performance:

The life-of-activity target is that 85% of revised land use plans in the focal municipalities will exclude all development from areas identified as sensitive wetlands. As of the midpoint of the activity, ten municipalities where the strategic approach had been implemented had released new land use plans, five of which met established best practices for zoning in sensitive wetland areas.

The mission is not sure how to interpret this evidence in the context of the learning question. Does this mean that the community engagement efforts were effective? Without a baseline assessment of the land use plans in effect at the start of the activity, they cannot be sure whether previous land use plans excluded development from sensitive wetland areas. It is possible that 50% of municipalities already had adequate zoning requirements in sensitive wetland areas. If so, there has been no improvement in plan quality. They also do not know whether strong community advocacy was associated with the zoning requirements in the new plans. It is possible that some of the plans with stronger zoning occurred in municipalities where community engagement was low or non-existent, which would contradict the assumption that community engagement influenced zoning outcomes.

What could the team have done differently?

The team realizes that they did not articulate their question in a way that clearly identified their information need. The team wants to know whether there is evidence that investing in community advocacy is an effective approach for influencing decision makers' land-use decisions, but the findings only tell them how many municipalities released plans with strong wetland protections. These performance indicators are an important component of their monitoring efforts, but on their own, they do not allow the team to make strong conclusions about effectiveness. The framing of their question did not make clear the hypothesis they wanted to test.

As the team approaches their final performance evaluation, the mission decides to include an objective that addresses this information need and modifies their question.

Objective: Assess the effectiveness of community advocacy as an approach for influencing protections for wetland areas in land use plans.

Question: In municipalities with insufficient wetland protections in land use plans, are stronger wetland protections more likely to be adopted in municipalities with high participation in community advocacy compared to those with low participation in community advocacy?

Given that there are no existing baseline data, the team will assess wetland protections in land use plans that were in effect at the onset of the activity. By limiting the population to municipalities with insufficient wetland protections in place at the start of the activity, the team will know that any revised plans containing strong wetland protections adopted stronger protections during the activity period. By comparing the strength of wetland protections in revised plans in municipalities where there was low participation in community advocacy and those where there was high participation in community advocacy, they can isolate participation as a possible explanation for any observed differences between these two groups. The team also carefully defines how they will decide what strong and weak protection and high and low participation mean in this context when designing this activity's MEL plan.

The final evaluation is completed and includes the following assessment:

Eighteen municipalities had insufficient wetland protections in their land use plans prior to the start of the activity. At the end of the activity, 11 of these municipalities were rated as having "high" participation in community advocacy groups supporting wetland protections and seven municipalities were rated as having "low" participation in community advocacy. Eight of 11 (73%) of "high" participation municipalities released revised plans including strong zoning in sensitive wetland areas compared to two of seven (29%) "low" participation municipalities. While sample sizes are low, these findings suggest that municipalities with high participation in community advocacy are more likely to include strong wetland protections in revised plans than municipalities with low participation. Participation in community advocacy is a possible cause for the observed differences in plans between the two groups. However, the evaluation cannot exclude other factors that may differ between the groups as also being contributors to improved wetland protections in the revised plans. An additional five municipalities included in the activity already had land use plans that restricted development in sensitive wetland areas and carried these requirements into revised plans released during the activity. In total, at the end of the activity, 15 out of 23 (65%) plans included zoning that excluded all development from sensitive wetland areas. These findings suggest that community advocacy may be an effective strategy for influencing land use planning decisions, but full achievement of the life-of-activity target remains dependent on the implementation success.

INTERPRETING EVIDENCE

Program managers should carefully review findings generated through research and monitoring, evaluation, and learning processes that are used in support of causal claims. Similar considerations apply when teams appraise findings from the evidence base (see "Practical Tips for Program Managers" in Section 5 of *Unit 2: Using Evidence*). The strongest arguments for causality incorporate three main lines of evidence (Trochim 2006, see Box 5 on page 25):

1. Time order: For the causal relationship to be valid, the presumed cause must precede or coincide with the observed effect.

2. Correlation: Changes in the observed effect must be associated with changes in the presumed cause.

3. Elimination of plausible alternative explanations: This condition is often the most difficult to establish.[7]

Ideally, teams will have evidence that shows statistical significance (to rule out random variation as the explanation for the findings) and will be able to incorporate different lines of mutually reinforcing evidence. These conditions may be fairly rare in practice, in which case teams should recognize the limitations of the evidence being used to inform decisions and seek alternative means of substantiating the findings when a decision requires certainty that X (and not some other factor) caused Y.

BOX 5: LIMITATIONS OF DIFFERENT DESIGNS FOR ESTABLISHING CAUSAL RELATIONSHIPS

Any analysis examining program effectiveness must account for the extent to which the data design addresses the three criteria used to establish causal relationships.

Descriptive designs provide the lowest support for causal claims. A descriptive design measures attributes on subjects in a single group but does not establish time order or association between variables. Assessing the number of animals illegally killed per year across multiple sites is an example of this type of design.

Before-and-after comparisons establish time order and association but confound causal factors with other factors that change over the same time period. For example, the number of animals illegally killed per year might be assessed before and after patrols were implemented across multiple sites.

Group designs disassociate the causal factor from other co-occurring factors across different study groups. Careful selection of comparison groups[8] can help mitigate selection bias arising from differences between the groups. For example, the number of animals illegally killed might be assessed across comparable sites that differ in patrol effort.

Experimental designs provide the greatest support for causal claims, but may not be technically feasible or ethically desirable for testing program assumptions because they require randomization. Random assignment of subjects to treatment and control groups minimizes the effects of confounding variables and selection bias, for example it might be possible to randomly assign patrol effort to sites in the group design above.

EXAMPLE 2: A PROGRAM MANAGER REVIEWS THE EVIDENCE SUPPORTING AN IMPLEMENTING PARTNER'S REQUEST TO EXPAND ITS CONSERVATION ENTERPRISES PROGRAM

An implementing partner recently completed a pause-and-reflect session reviewing their year three activity outcomes. The activity is partially funded with biodiversity funds and focuses on reducing unsustainable timber extraction in a tropical forest system. The team reviewed several lines of evidence that appear to support the conclusion that additional income from an ecotourism enterprise allows participants to reduce their reliance on timber extraction (a traditional source of income). They are requesting a modification to their contract that would allow them to divert resources from other strategic approaches in order to expand the ecotourism enterprise. The program officer reviews the evidence provided by the implementation team, paying careful attention to how well it supports the team's conclusion.

Situation A: The team has data from before and after the ecotourism enterprise was established showing that the average amount of timber extracted per year decreases among participants who receive income from the enterprise.

The findings establish time order because the decrease in timber extraction occurred after participants started receiving additional income. The before-and-after comparison also establishes a correlation between timber extraction and income from the conservation enterprise. However, the study design cannot rule out alternative explanations that may have changed over the same time period. Perhaps increased enforcement of illegal harvest was put in place at the same time the ecotourism enterprise was established, so the observed decline in forest use might not be related to income from the conservation enterprise at all. The program officer might consider asking the team to provide additional evidence showing that the trend among enterprise participants differs from the broader population of forest resource users in the area.

Situation B: The team has data showing that participants in the ecotourism enterprise extract fewer timber products per year than non-participants.

Here the data show a significant correlation between participation in the enterprise and timber extraction. Several plausible explanations are consistent with the observed correlations, and they cannot be ruled out using only these findings. For example, people who chose to participate in the enterprise may have been minimally engaged in timber extraction to begin with. If so, the difference between the groups is pre-existing and likely due to a selection bias rather than participation in the enterprise. The program officer might consider asking the team for baseline data that confirms that participants extracted similar amounts of timber relative to the broader population of resource users prior to the onset of the activity.

Situation C: The team has data from before and after the ecotourism enterprise was established showing that the average amount of timber extracted per year decreases among participants who receive income from the enterprise. However, the magnitude of the observed decrease is small and statistically insignificant.

This situation is the same as that in Situation A, but with the added limitation that the magnitude of the change is not sufficiently large to rule out that it was produced by chance. Even though the data show a change in direction that is consistent with the team's hypothesis, they are not strong enough for the team to make a definitive conclusion about the effectiveness of their strategic approach. The program manager might ask the team to expand their data collection efforts across a larger sample or to increase the length of time through which they will collect data.

Situation D: The team has data from before and after the ecotourism enterprise was established showing that the average amount of timber extracted per year decreases among participants who receive income from the enterprise. The data also show that forest users not participating in the enterprise did not change their practices over the same time period. The team interviewed a representative sample of participants, and the interviewees singled out the additional income as the primary motivator for the changes in forest use.

In this case the data support the hypothesis that the expected behavior change started after the strategic approach was in place. An appropriate comparator allows for assessing the magnitude and direction of the change, and an independent line of evidence helps the team infer that income from the ecotourism enterprise was the main driver of the observed changes. This scenario provides the strongest support for the team's hypothesis that the ecotourism enterprise is an effective way to reduce the amount of timber being extracted by individuals in the program context. These findings provide the strongest evidence for the team's request.

6. SUMMARY OF KEY CONCEPTS

- Better evidence leads to better programming decisions. Once a team is aware of information needs and the points at which they are likely to be identified in the Program Cycle, they can consider generating evidence to address them.

- There are three general approaches that teams can use to generate evidence about the validity of program assumptions:

  (1) Commissioning research through the procurement of activities,

  (2) Designing evaluation questions to strengthen understanding of the theory of change and its implementation, and

  (3) Collecting relevant data as part of monitoring of implemented activities.

- Ensuring that the team's questions are clearly articulated and feasibly researchable will increase the likelihood that relevant and credible evidence will be generated.

- The strongest arguments for causality incorporate three main lines of evidence: time order, correlation, and elimination of plausible alternative explanations. Ideally, teams are able to generate evidence that shows statistically significant differences or incorporate different lines of mutually reinforcing evidence.

7. FURTHER READING

Setting research priorities:

USAID Biodiversity and Development Research Agenda (USAID 2015). This research agenda defines and prioritizes the most critical research needed in the area of biodiversity conservation in support of USAID's conservation and development objectives. Annex A describes the approach and methodology used for identifying and prioritizing research topics in the agenda.

Framing questions and selecting a research approach:

Research Questions and Methodologies for a Biodiversity and Development Research Agenda (USAID 2016b). This USAID brief discusses how to formulate a research question that can be operationalized. It also reviews common research methodologies that can be used to support evidence-based programming.

Generating evidence:

Research Methods Knowledge Base (Trochim 2006). This online resource offers an easily navigated and comprehensive introduction to social research methods. It covers the entire research process from formulating research questions, to research design, and data analysis. Particular topics of interest include an explanation of types of questions and the section on design which includes a discussion on the strengths and weaknesses of different designs.

ENDNOTES

1 Box 2 in *Unit 1: Understanding an Evidence-Based Approach* includes a representation of three components of program success.

2 Political Economy Analysis is a field-research methodology used to explore the causes of a development or governance issue or a problem in implementation. See *Using Political Economy Analysis for Biodiversity Conservation Planning* for a case study applied to the biodiversity sector.

3 *Foreign Assistance Act Sections 118/119 Tropical Forest and Biodiversity Analysis: Best Practices Guide* (Martino et al. 2017) provides further information about the use of the Biodiversity and Tropical Forestry Assessments in biodiversity programming.

4 Including a comparison group in the question that addresses the causal claim in the theory of change for conservation enterprises (see page 18) would help rule out alternative explanations that might affect household income for all stakeholders regardless of whether they participated in a conservation enterprise or not.

5 The components used to articulate testable questions in biodiversity programming are similar to those used in other areas of evidence-based practice, see Davies (2011).

6 Benchmarking is a method of evaluation comparing performance against a standard. The approach presented in *Biodiversity How-To Guide 3: Defining Outcomes & Indicators in USAID Biodiversity Programming* uses outcome statements as benchmarks against which to compare indicators.

7 "Designing Designs for Research" in Trochim (2006) provides further discussion of options for minimizing alternative explanations for hypothesized cause-effect relationships.

8 A USAID technical note on impact evaluation provides a useful discussion on selecting comparison groups that also applies to other types of studies.

GLOSSARY

Assumption: Used in *Evidence in Action* to refer to the logical connections between drivers, threats, and the status of biodiversity focal interests in a problem analysis or those that underlie anticipated results articulated in a program's theory of change.

Biodiversity focal interests: The species, habitats, and/or ecosystems that a program is working to conserve.

Claim: In *Evidence in Action*, refers to a statement that describes what would be observed if a given program assumption is true. When the validity of a claim is uncertain, it can be considered a hypothesis and is used to direct evidence generation by way of asking a testable question.

Confounding: Occurs when one or more outside factors co-varies with a presumed cause, making it difficult to establish the true cause of an observed effect.

Counterfactual: In the context of evaluating program effectiveness, a counterfactual refers to a group or site that was not exposed to a strategic approach.

Effectiveness: The degree to which an implemented project or activity achieves intended outcomes. Understanding the effectiveness of a strategic approach involves testing the assumptions that underlie a program's design.

Evidence: The body of facts or information that serve as the basis for programmatic and strategic decision making in the Program Cycle (ADS Chapter 201, page 145). Used in *Evidence in Action* to refer to (1) individual findings or pieces of information used to help make a decision or support a conclusion; and (2) the body of findings or information providing support for (or countering) a belief or claim related to effectiveness or attribution.

Evidence-based approach: The conscientious, explicit, and judicious use of current, best evidence in program decisions. An evidence-based approach encompasses identification, use, and generation of evidence to increase program effectiveness.

External validity: The extent to which the findings from one study can be applied to other contexts.

Grey literature: Documents and other materials produced outside of commercial or academic publishing and distribution channels, including government agencies, universities, corporations, non-governmental organizations, societies, and other professional organizations.

Hypothesis: An explanation for a phenomenon that can be verified by investigation or methodological observation.

Program (and Programming): Used in *Evidence in Action* as a general term to encompass USAID project and activity levels.

Selection bias: An artificial skewing of the results caused by non-random selection of individuals, groups, or data.

Situation model: A graphic representation of a context or problem analysis (often called a conceptual model).

Strategic approach: A set of actions with a common focus that work together to address specific threats, drivers, and/or opportunities in order to achieve a set of desired results.

Testable question: A question formulated to generate evidence that can support or refute a specific hypothesis.

LITERATURE CITED

Davies, K. S. 2011. Formulating the evidence based practice question: A review of the frameworks. Evidence Based Library and Information Practice 6:75–80. https://doi.org/10.18438/B8WS5N

Martino, R., K. Menczer, and H. Kushnir. 2017. Foreign Assistance Act 118/119 Tropical Forest and Biodiversity Analysis: Best Practices Guide. USAID Bureau for Economic Growth, Education, and Environment Office of Forestry and Biodiversity. Washington, DC. http://pdf.usaid.gov/pdf_docs/PA00MKS3.pdf

Sterling, E. J., E. Betley, A. Gomez, A. Sigouin, C. Malone, M. Blair, G. Cullman, Chris Filardi, K. Landrigan, K. Roberts, and A. L. Porzecanski. 2016. Stakeholder Engagement for Biodiversity Conservation Goals: Assessing the Status of the Evidence. USAID Bureau for Economic Growth, Education, and Environment Office of Forestry and Biodiversity. http://pdf.usaid.gov/pdf_docs/pa00m2m6.pdf

Trochim, W.M. 2006. The Research Methods Knowledge Base, 2nd Edition. http://www.socialresearchmethods.net/kb/ (Version current as of October 20, 2006).

USAID. 2013. Impact Evaluations. USAID Bureau for Policy, Planning, and Learning. Washington, DC. https://www.usaid.gov/sites/default/files/documents/1870/IE_Technical_Note_2013_0903_Final.pdf.

USAID. 2015. Biodiversity and Development Research Agenda. USAID Bureau for Economic Growth, Education, and the Environment Office of Forestry and Biodiversity. Washington, DC. http://pdf.usaid.gov/pdf_docs/pa00kb5x.pdf

USAID. 2016a. Tips for Evaluating Good Evaluation Questions (for Performance Evaluations). USAID Bureau for Policy, Planning, and Learning. Washington, DC. https://usaidlearninglab.org/sites/default/files/resource/files/tips_for_developing_good_evaluation_questions_2016.pdf

USAID. 2016b. Research Questions and Methodologies for a Biodiversity and Development Research Agenda. USAID Bureau for Economic Growth, Education, and the Environment Office of Forestry and Biodiversity. Washington, DC. http://pdf.usaid.gov/pdf_docs/pa00kxmt.pdf

U.S. Agency for International Development
1300 Pennsylvania Avenue, NW
Washington, D.C. 20523
Tel: (202) 712-0000
Fax: (202) 216-3524
www.usaid.gov/biodiversity