Africare
M&E Toolkit

Tools and Methods
For a Responsive and Integrated Monitoring and Evaluation System

Prepared by
Madeleine Gauthier, Consultant
S. Joshua Volle, DrPH,
Director of M&E, Africare
October 2014

Improving lives, Building futures
# CONTENTS

**INTRODUCTION**  
What do we mean by M&E and why everyone should be concerned  
How to use this toolkit  

**MODULE 1  FORMATIVE ASSESSMENTS**  
1.1 Formative Needs Assessment  
1.2 What are the basic steps to conducting a Formative Research?  
1.3 Formative Needs Assessment Methods  
1.4 Beneficiary Analysis  

**MODULE 2  MONITORING**  
2.1 Monitoring  
2.2 Monitoring questions to be answered  
2.3 Choosing indicators  
2.4 Data Collection  
2.5 Data Base  
2.6 Supervision and Data Quality Assessment  

**MODULE 3  EVALUATION**  
3.1 What is an Evaluation  
3.2 Types of Evaluations  
3.3 Factors to consider when deciding on an Evaluation  
3.4 The Counterfactual: the key to attribution  
3.5 Putting together a protocol and SOW for an Evaluation  

**MODULE 4  PROJECT INDICATORS**  
4.1 Selecting, Testing, and Using Indicators  
4.2 Custom vs. Standard or Common Indicators  
4.3 Indicator Level  
4.4 Characteristics of good performance indicators  

**MODULE 5  USE OF THE RESULTING M&E INFORMATION**  
5.1 Project self-assessment  
5.1.1 Process Evaluation  
5.1.2 Outcome Evaluation  
5.1.3 Quality assessments  
5.2 Project expansion and replication  
5.3 Marketing our successes  
5.4 Reporting  

**MODULE 6  THE PRE-AWARD PHASE**  
6.1 Articulating a results statement  
6.2 Constructing the project logical framework  
6.3 Pre-Award M&E Plan/proposal  
6.4 Budgeting for M&E  

**MODULE 7  THE PERFORMANCE MONITORING PLAN OR M&E PLAN**  
7.1 Introduction to the Performance Management Plan or M&E Plan  
7.2 Components of a PMP or M&E Plan  
7.3 Performance Indicator Reference Sheet
MODULE 8 PUTTING THE PMP TO WORK 46
8.1 THE INDICATOR PERFORMANCE TRACKING TABLE 47
8.2 DEVELOPING OTHER STANDARD FORMS 47
8.2.1 Data Flow Diagram 47
8.2.2 Data collection forms 49
8.2.3 Reporting forms 49
8.3 DATA MANAGEMENT SYSTEMS 49
8.4 DATA QUALITY MANAGEMENT PROTOCOLS 50
8.4.1 Data quality insurance plan 54
8.4.2 Internal data quality assessments 54
8.4.3 External data quality assessments 55
8.4.4 The feedback loop 55
8.5 CALENDAR FOR REVIEW OF THE IPTT AND FEEDBACK LOOP 55
8.6 PERIODIC REVIEWS OF THE M&E PLAN 56
MODULE 9 METHODS AND TOOLS FOR DATA COLLECTION 57
9.1 TYPES OF DATA 58
9.1.1 Primary versus secondary data 58
9.1.2 Quantitative versus qualitative data 58
9.2 METHODS OF DATA COLLECTION 59
9.3 QUALITATIVE METHODS 59
9.3.1 Interviews 60
9.3.2 Focus Groups 60
9.3.3 Direct observation 61
9.3.4 Mystery Clients 61
9.3.5 Hearsay Ethnography 62
9.3.6 What to include in a qualitative Scope of Work 62
9.4 QUANTITATIVE DATA COLLECTION TOOLS 62
9.5 SAMPLING 63
9.5.1 Non-probability sampling 64
9.5.2 Probability sampling 64
9.5.3 Cluster sampling strategy 65
9.5.4 Sample size 65
9.5.5 Stratification 66
9.6 DATA QUALITY 66
9.7 DATA ANALYSIS 67
9.8 WHAT TO INCLUDE IN A SURVEY SCOPE OF WORK (SOW) 68
MODULE 10 ANNEX 71
10.1 GLOSSARY 72
10.2 SAMPLE SIZE CALCULATIONS 75
10.3 SURVEY RESPONDENT CONSENT FORMS 76
10.4 QUALITATIVE METHODS GUIDANCE 81
10.5 INDICATOR REFERENCE SHEET FOR THE PMP 86
10.6 ILLUSTRATIVE INDICATOR TABLE FOR THE PMP 88
10.7 ADVANTAGES AND DISADVANTAGES OF VARIOUS DATA COLLECTION METHODS 89
10.8 EVALUATION RESEARCH PROTOCOL TEMPLATE 92
INTRODUCTION

The overall purpose for this toolkit is to move Africare’s monitoring and evaluation efforts in the same direction. While it is important for all M&E systems to be flexible and responsive to the local context and demands it is also possible for each system to be based on and developed out of the same general set of methods, beliefs and assumptions. A large part of this general similarity falls into the arena of language and how we talk about what we are doing. The intention for this M&E Toolkit is to lay out that common language to ensure that we are all on the same page and able to discuss our M&E activities without adding confusion.

In addition, this Toolkit is a part of the overall quality assurance system for Africare’s M&E activities. Throughout the Toolkit you will find suggestions and sometimes requirements for how to conduct the M&E business of Africare. These suggested and required methods and approaches support Africare in the development of rigorous, efficient, useful and valuable M&E systems wherever we are working.

At the end of the day, following these guidelines and agreed upon methods will support Africare’s ability to tell the story of Africare and the terrific work that is being carried out in order to improve the lives of the people of the African continent.

What do we mean by M&E and why everyone should be concerned

The purpose of Monitoring and Evaluation (M&E) is gather information and data to tell our story. Data that is collected and left sitting on a computer or printed out and left on a shelf is a huge waste of human and financial resources and is not M&E. When we develop an M&E system we are creating a system of information and communication that allows us to describe the “who”, “what”, “where”, “when”, and “how much” of each project and program. In doing this, we will have what we need to describe our process, identify results, report to funders, market our good work, and have sufficient documentation to replicate successful projects in new areas.

A primary use of M&E data is to measure the extent to which the proposed interventions are indeed leading to the desired outcomes; to track the progress of implementation. From the implementation point of view, it is important to track the result path against a timeline and resource plan in order to make adjustment to the project if things are not unfolding as planned or results are not manifesting as had been theorized during the planning stage. So the information is used to tell how well we are doing and whether or not changes need to be made.

It is also the basis of the information used to report to headquarters and project funders so that we can show where the money is being spent and ensure them that we are carrying out the funded activities or be able to explain why they are not being carried out as planned. M&E data is also a good source of data that can be used to build a body of objective, documented evidence on implementation strategies and techniques and develop best practices.
By monitoring, we usually mean the periodic, systematic collection of information during the implementation of a project through a pre-defined set of indicators and targets. Evaluation refers to a discreet collection of information to assess the relevance and degree of achievements of certain hypotheses made by a project, such as meeting expected outcomes or using specific implementation strategies. Monitoring is the responsibility of the implementation team, while different stakeholders (funding agencies, governments, development professionals) share the evaluation challenge. To be credible, an independent party should conduct an evaluation, although the implementation team still bears the responsibility towards the availability and comprehensiveness of the implementation data.

The majority of the funding agencies require that some kind of M&E plan be developed as a part of any project or program. The core of the M&E plan is developed during the proposal stage and expanded upon once an award has been made. In any comprehensive program there are a number of data points that will eventually all come together to form your M&E system. Later in this tool kit we will address the M&E Plan which is a description of each piece of this system. The system consists of:

1. Formative Assessments
   a. Situation analysis
   b. Communication analysis
   c. Can include the baseline data

2. Monitoring program activities
   a. Routine process data
      i. Process — Set of activities in which program resources (human and financial) are used to achieve the results expected from the program (e.g., number of workshops or number of training sessions).
      ii. Outputs — Immediate results obtained by the program through the execution of activities (e.g., number of commodities distributed, number of staff trained, number of people reached, or number of people served).

3. Program quality

4. Process Evaluation
   a. Program Targets
      i. To assess the extent to which objectives have been met. It is the process of determining the effectiveness of a program or a project.
   b. Routine process data
      i. Compared to targets

5. Program outcome and impact Evaluation
   a. To assess the changes in the target group (e.g., changes in risk behavior)
      i. Outcomes—Short-term or intermediate results obtained by the program through the execution of activities
      ii. Impact—Long-term effects (e.g., changes in health status). This can be through special studies with wide district, regional, or national coverage
   b. Baseline data
   c. Mid-term follow-up
   d. End-line follow-up

**Comprehensive Monitoring and Evaluation Framework**
Types of Monitoring and Evaluation:

<table>
<thead>
<tr>
<th>Formative Assessments and Research (concept and design)</th>
<th>Monitoring (monitoring procedures, and outputs; assessing quality)</th>
<th>Evaluation (assessing outcome and Impact)</th>
<th>Cost-Effectiveness Analysis (including sustainability issues)</th>
</tr>
</thead>
</table>

Questions Answered by the Different Types of Monitoring and Evaluation:

- Is an intervention needed?
- Who needs it?
- What is the best design?
- To what extent are planned activities actually realized?
- How well are the services provided?
- What outcomes are observed?
- What do they mean?
- Does the program make a difference?
- Should program priorities be changed or expanded?
- Should resources be reallocated?

How to use this toolkit

The challenge of this toolkit is to provide guidance and develop standards for an agency that is highly decentralized and must remain operationally flexible in order to respond to the evolving needs and constraints of its beneficiary population and sources of funding. An M&E system however, almost by definition, imposes a certain amount of rigidity and standardization of procedures can meet resistance if it is seen as impeding flexibility.

Another challenge of the toolkit is to contend with the dual purposes of M&E, that of serving external reporting requirements and as an internal management tool. While everyone wishes to see M&E first and foremost as a management tool, its support function to reporting typically takes over the time and resources allocated to M&E.

Hopefully, this toolkit offers both a “flexible” document to learn more about M&E, what it can do (and can’t do) and how it is best undertaken with a learning and adapting lens. It is intended to be a resource where one can look up (and be reminded of) M&E definitions and techniques as well as find useful references to more in-depth discussions of concepts and techniques. Most references have links to where the source document can be found on the Internet.

The toolkit is arranged in modules. To the extent that Africare has established procedures, such as with the “Problem Analysis” framework followed in project design or the IPTT in the 3PR process, the toolkit will focus on these tools. Recognized best practices are outlined and additional references are provided. The importance of the U.S. Government as a donor, and of USAID in particular, calls for emphasizing USAID guidelines where relevant. Examples are included as often as possible as a way to simplify, focus, or reinforce the reader’s understanding of relatively new concepts or techniques. These examples are drawn to the extent possible from Africare’s experiences and lessons learned. References and recommended readings are listed at the end of sections or modules, with a link of where these can be found on the Internet. We recommend downloading these references as appropriate and keeping all resources together so that they can be easily and regularly consulted. Users are encouraged also to add and share other examples and references that could help in making this toolkit as useful as possible.
MODULE 1  FORMATIVE ASSESSMENTS

PURPOSE OF THE MODULE
This first module is designed to provide the basic information needed to carry out a formative assessment. A formative assessment is an investigative process that is carried out prior to the development of an intervention or activity. The aim of the investigation is to discover key issues and elements of the community that lead to a particular development/health outcome or concern.

The key concepts to take away from this are:
1. it occurs prior to the development of the intervention; and
2. is a data gathering investigative activity

The information that you obtain during this activity will lead to increased community involvement and provide a solid basis for decisions that you will make in designing the best intervention possible. With this information you will be able to identify the exact nature of the concern, what behaviors are taking place that put people at risk for the concern, which members of the community are most at risk, what are the determinants to the behavior(s), what resources are available and what has been tried before. All of this should be taken into consideration when developing your intervention and the results of your formative assessment/research should lead directly into specific actions and interventions.

A large part of the research will also contribute to the baseline information for your long-term program evaluation.

The bottom line is that we conduct formative assessments/research in order to:
• Clearly understand the development/health concern
• Begin to engage the community in the concern
• Obtain all the necessary information to design an intervention, and
• Begin the program Monitoring and Evaluation Process

CONTENT OF THE MODULE

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Formative Needs Assessment</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>What are the basic steps to conducting a Formative Research?</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Formative Needs Assessment Methods</td>
<td>2</td>
</tr>
<tr>
<td>1.4</td>
<td>Beneficiary Analysis</td>
<td>3</td>
</tr>
</tbody>
</table>
1.1 Formative Needs Assessment

Not only is this an important tool for program designers and developers but at the same time it is the cornerstone of any solid M&E system. In an ideal setting the Formative Needs Assessment would be conducted during the planning (or re-planning) stage of a program to identify community needs, issues, experiences, beliefs, potential partners and potential barriers before a program is widely implemented. This is the point where flexibility is greatest and program sponsors have more freedom to make decisions about how to proceed with implementation. More often than not, however, there is no time or funding to carry out these assessments prior to project development. In that case we rely on what has already been done in previous projects or we build it into the start-up of the project.

During a Formative Needs Assessment, the following issues are explored:

a) Identifying the need for interventions
b) Segmenting your target audiences
c) Mapping out the current resources and services
d) Identifying channels of communication, cultural and through media
e) Defining realistic goals and objectives for interventions
f) Identifying feasible program strategies
g) Setting program targets

Use the Formative Needs Assessment as an exploratory tool as well as a way to help project managers adjust objectives to changing situations. You can also use it to identify unacceptable or ineffective intervention approaches, designs, and concepts.

1.2 What are the basic steps to conducting a Formative Research?

Before beginning your formative research you must first plan out all of the steps and activities that you will need to carry out. The following is an outline of the basic steps to follow when conducting any community assessment. All of the issues and questions discussed in the previous section of this module can be gathered through this process. The following section goes into detail on each of these steps.

1. Form a community needs assessment committee
2. Decide what specific information you will want to collect
3. Decide who your target is for data collection: who will you want to talk with and where are they located
4. Decide on the methodology to best fit the situation and available resources
5. Decide on the timetable so that the information gathered is timely and relevant to the program design
6. Implement the assessment using the selected tool(s)
7. Use report to design and inform the development of the intervention(s)

1.3 Formative Needs Assessment Methods

Many methods are useful during this process. Typically we triangulate multiple methods in order to provide the clearest picture of what the current status and situation is related to the development/health issue we are concerned with. The methods include both
qualitative and quantitative methods and are normally carried out as rapid assessments in order to ensure that the information is timely and the costs are kept to a minimum.

Some methods used for formative assessments are:

a) **Reviews of existing information**: While this should always be carried out it is particularly important to look at existing information if time is limited or the project funder does not allow any time for assessments prior to design. This information may come from published literature, Africare country program experience, or partner experiences that are related to the area of concern. It is important to keep aware of the need for the information to be directly related to the target population, using information on results in Senegal do not necessarily translate into the same response/results in Mozambique.

b) **Focus group discussions**: FGDs are not group meetings. There are specific guidelines and structures to FGDs that must be adhered to (See Annex 9.4 for guidance). Used properly they can yield a great deal of cultural relevant information and insight into what is happening with this target group.

c) **Individual in-depth interviews**: IDIs allow you to take a topic of interest and dive deeply into it with experts on the topic. These experts may be community members who have the experience we are looking to work with or they may be gate keepers or other kinds of leaders in the community. They may also be professionals with experience working on the topic of concern.

d) **Participant observations**: These are a little trickier than other methodologies because they may require the observer to participate in risky situations or behaviors, not recommended. One specific methodology to try out is called “Hearsay Ethnography” (See Annex 9.4). This uses community members as ethnographers who have been trained to listen and notice things as they go about their normal day. It takes a good bit of training and at least a few weeks of observations by the participants before rich information comes out but it eliminates the community feeling like an outsider is sitting among them judging and watching.

e) **Surveys with structured questionnaires**: Such surveys can take up a good bit of resource but if the time and money is available they can become invaluable. When these are done properly they can actually turn into your baseline data that is then used at mid-line or end-line to carry out the outcome evaluation. If that is the intention then the survey must be carried out using proper sampling and data collection methods that will then be repeated later in the life of the project.

1.4 **Beneficiary Analysis**

Beneficiary analysis does not refer to a specific standard methodology but it is a very important part of the formative assessment. What it refers to is an analysis of available data that can bring light to the distributional effect of proposed interventions; will some beneficiaries be affected more or less than others by the proposed intervention? For example, while data on income distribution are notoriously scarce and may not be transferrable to a project-level area (province, district), it is useful to look at what is available and try to understand what are the income characteristics of the targeted population and where do they appear to be situated in the income distribution scale. A project that seeks to expand export crops may hold a lot of promise in terms of increased
income in a specific district, but it may target farm households relatively well-off or an area better endowed than other parts of the country. Another interest for conducting a beneficiary analysis may be to investigate income generating activities of households. A project may directly or indirectly affect how households choose to allocate their time between different income generating activities, or could affect members of household differently in their access to income in ways that were not envisaged.

These kinds of analysis can help in justifying the project by demonstrating desirable impact or that undesirable ones can be avoided and in any case, it can highlight trends and risks to be monitored during implementation.

The main limitation of most of these methods for carrying out the Formative Needs Assessment is their inability to be generalized to other projects. They will be invaluable to the project being designed but of limited use to others, with the exception of the surveys if carried out following standard research protocols.

The following image shows that the results of the assessment are used in program/project design as well as in establishing the basis for the M&E system. Notice also that the various levels of the M&E system, the pieces to be monitored and evaluated, directly correspond to the various levels of the project’s objectives and goal.
MODULE 2 MONITORING

PURPOSE OF THE MODULE
In any comprehensive program there are a number of data points that will eventually all come together to form your M&E system. In the previous module we covered the preliminary work carried out through the formative assessment or research. In this module we turn to the routine monitoring data and the role that it plays in project implementation, re-design and even evaluating the process of the project/program.

Monitoring data are the nuts and bolts of what the project is doing. These are the base data that are routinely collected to measure activities that the staff or volunteers are carrying out on a daily or weekly basis and are aggregated upwards to total amount of these actions taken in a month, quarter, year or over the life of a project. Another important measure of this routine data is the “reach” of the project; how many of our target population were actually engaged during any particular period of time.

Routine monitoring requires clear and easy to use data collection tools, a data base system that allows for aggregation of data, training of front line workers on data collection/recording of their activities, supervision of the front line workers, and periodic data quality assessments.

CONTENT OF THE MODULE

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>MONITORING</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>MONITORING QUESTIONS TO BE ANSWERED</td>
<td>6</td>
</tr>
<tr>
<td>2.3</td>
<td>CHOOSING INDICATORS</td>
<td>6</td>
</tr>
<tr>
<td>2.4</td>
<td>DATA COLLECTION</td>
<td>7</td>
</tr>
<tr>
<td>2.5</td>
<td>DATA BASE</td>
<td>7</td>
</tr>
<tr>
<td>2.6</td>
<td>SUPERVISION AND DATA QUALITY ASSESSMENT</td>
<td>8</td>
</tr>
</tbody>
</table>
2.1 Monitoring

Monitoring is the routine process of data collection and measurement of progress toward program objectives. It is about your process or procedures. There are two main categories of information required in a monitoring system, Inputs and Outputs. What you measure in these categories is always dependent on what your process objectives are or what the set of activities within general objectives are determined to be.

**Inputs**: Set of resources (human and financial) through which the program is carried out and are used to achieve the results expected from the program (e.g., number of micronutrient packets available or number of people trained).

**Outputs** – Immediate results obtained by the program through the execution of activities and measurements for how the population interacts with your program (e.g., number of commodities distributed, number of people reached, or number of people served).

2.2 Monitoring questions to be answered

Monitoring addresses the following questions:

- To what extent are planned activities actually realized? Are we making progress toward achieving our objectives?
- What services are provided, to whom, when, how often, for how long, and in what context?
- What is the quality of the services provided?
- What is the cost per unit service?
- Is the program undertaken consistently with the design or implementation plan?
- How well is the program reaching the intended target group?

2.3 Choosing indicators

Quite often indicators are predetermined by the donor or some other internationally recognized body. Always look to make sure that you include relevant indicators from the project donor and then add other indicators that come from internationally recognized bodies. Once you have these basic indicators selected your task now is to ensure that the key activities for each objective or result are being measured. This may mean developing your own set of indicators to capture the richness of the activities and efforts of the staff. Keep in mind that at this stage, indicators are to be selected or designed to measure actions taken by the program staff.

The intention here is to capture/measure the activities that make up the project and be able to concisely describe all the work that goes into a project so that the “model” can be replicated if it is found to be successful.
Bottom line, your indicators should each be:

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational</td>
<td>Should be measurable or quantifiable using tested definitions and reference standards.</td>
</tr>
<tr>
<td>Reliable</td>
<td>Should produce the same results when used more than once to measure the same condition or event.</td>
</tr>
<tr>
<td>Valid</td>
<td>Should measure the condition or event it is intended to measure.</td>
</tr>
<tr>
<td>Specific</td>
<td>Should measure only this condition or event.</td>
</tr>
<tr>
<td>Sensitive</td>
<td>Should reflect changes in the state of the condition or event under observation.</td>
</tr>
<tr>
<td>Affordable</td>
<td>Should represent reasonable measurement costs.</td>
</tr>
<tr>
<td>Feasible</td>
<td>Should be able to be carried out in the proposed data collection system</td>
</tr>
</tbody>
</table>

Examples of indicators for SBCC process are:
- # community design meetings were held in the past year
- % of trained peer educators providing accurate information 6 months after training
- # of peer educators trained, by sex
- # of beneficiaries reached by peer educators, by sex
- # of radio spots developed and pilot tested
- # of materials developed
- # of materials distributed

2.4 Data Collection
The data at this level is considered routine data that the program automatically collects. Data collection instruments must be created to ensure consistent interpretation of the indicators and consistent collection of information from one person to the next. All data collection tools for routine monitoring purposes must be pilot tested with the staff that will be responsible for using them. This should be built into the development process as well as a part of the staff/volunteers’ training. Their input will be invaluable in making sure that the forms are clear and measure what is required. It also increases the likelihood that they will (1) understand the tools and (2) buy into the use of the tools. At the same time, don’t allow staff/volunteer input to change the form so much that it is no longer capturing what is required for the routine reporting and supervision.

2.5 Data Base
Any good spreadsheet will do for this purpose. Most projects prefer to keep the routine data stored in Excel or Access. These are easy to use and come with most computer software packages. There are several key things to keep in mind when selecting the data base that you will use for any one project:
- Ability to house all of the data
- Ability to aggregate the data over reporting periods
- Ability to communicate with other systems
- Ease of use
- Cost
Most of Africare relies on Excel because it is easy to use and comes with the computers. But also a country based excel data base can be used to aggregate the data across reporting periods and then communicate directly with the PPR. So it is highly recommended that all projects store routine data in Excel.

The data base should be set up once the final list of indicators has been agreed up with the funder and at the same time as the data recording tools are being developed for the front line workers. These two must directly relate to each other since the data reporting tools will be used to enter the data into the data base.

Pay close attention to the indicators that must be reported on. Some of them will be percentages. When that is the case you need to ensure that the data collection/reporting tools collect the numerators and denominators when needed. Sometimes the denominator comes from other project records and can be set up in the data base as a constant number. For example, if you know that in the project region there are 52 pastoralists then you can enter that as your constant denominator in the data base. The numerator will come from the weekly tools reporting number of pastoralists reached with animal vaccination information.

An ongoing concern at this level of monitoring is the counting of beneficiaries and how to avoid double counting, especially when we are aggregating the numbers over time. One approach is to have the frontline staff reporting on their weekly reports how many of their contacts were first time contacts and how many were second or third contacts during that reporting period. Then in the data base the first time contacts for each week can be aggregated across a month or a quarter to show the total number of first time contacts, a good estimate of the reach of the project for that month and quarter. Another approach is to assign ID numbers to each beneficiary and have the weekly or monthly reports come in based on those ID numbers. This can be very complicated for large projects but provides more accurate information when it is possible.

2.6 Supervision and Data Quality Assessment
The results can only be as good as the quality of the data coming into the system. The first level of quality data is the creation of clear and easy to use data collection/reporting tools. The second level of quality data is the provision of sufficient training on those data collection/reporting tools. And the third level of quality data is supervision of the use of the data collection/reporting tools.

There must be ongoing and routine supervision of the data collection. This basically relies on a highly skilled supervisor reviewing the data collection/reporting forms each day or week for clarity, completeness and accuracy and bringing any concerns to the attention of the specific staff/volunteer responsible for the forms in question. It also entails period re-training and directly observing the staff/volunteers as they are filling out the forms.

Data Quality Assessment (DQA) is a periodic, usually annual, activity carried out to look at the source data and compare that source data to what has been reported to the funder. It actually begins with the data reported to the funder and follows the system backwards to ensure that the numbers reported at the end of the system are in alignment with the
data at each stop along the M&E system. By following the data back to the source we are able to find out where, if at all, there were problems with the data. Once a stage is identified to have problems we can then develop plans to alleviate the problems and improve the M&E system so that we are reporting high quality data.

In summary, good quality data is based on:

- Development of quality data collection/reporting instruments
- Input from frontline users of the instruments
- Training of frontline users
- Supervision of frontline users as they report their activities
- Re-training/support for frontline users
- Annual DQA
- Addressing issues brought up during annual DQA
MODULE 3  EVALUATION

PURPOSE OF THE MODULE

This module presents a brief overview of concepts and techniques used in evaluation. There has been a renewed interest in evaluation in the past few years, in part as a way to investigate questions raised by a widening community of social scientists, development professionals, politicians, and taxpayers concerned about aid effectiveness. An evaluation has two functions—learning and accountability—which should be considered mutually reinforcing. An evaluation should be as objective as possible and should be collaborative and transparent in its purpose and method.

Even if Africare has a limited mandate to conduct evaluations, it has a crucial role in supporting and facilitating evaluations done by other organizations. This is to ensure that the findings are relevant and useful and contribute to the shared objective of improving aid effectiveness.

CONTENT OF THE MODULE

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 WHAT IS AN EVALUATION</td>
<td>11</td>
</tr>
<tr>
<td>3.2 TYPES OF EVALUATIONS</td>
<td>11</td>
</tr>
<tr>
<td>3.3 FACTORS TO CONSIDER WHEN DECIDING ON AN EVALUATION</td>
<td>13</td>
</tr>
<tr>
<td>3.4 THE COUNTERFACTUAL: THE KEY TO ATTRIBUTION</td>
<td>19</td>
</tr>
<tr>
<td>3.5 PUTTING TOGETHER A PROTOCOL AND SOW FOR AN EVALUATION</td>
<td>20</td>
</tr>
</tbody>
</table>
3.1 What Is an Evaluation

When we discuss monitoring, we mean the periodic and systematic collection of information during the implementation of a project using a set of pre-defined indicators and scales (targets). Evaluation, on the other hand, refers to a “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.”

International organizations have arrived at the consensus that there are three primary reasons for conducting evaluations:

- Program improvement, using evaluation results as feedback to program implementers to make the program function more effectively and efficiently.
- Program accountability and transparency, so stakeholders and funders alike are aware of the progress of the program.
- Program scale-up, disseminating results to help stakeholders and partners better understand what the program has accomplished and to replicate similar approaches in the future.

Ultimately, evaluations are conducted in order to answer specific questions and inform resource allocation and policy decision about future projects. An evaluation can tackle a project-wide question such as whether the implementation went as planned and resources were used effectively. It can also focus on a specific issue, such as explaining an expected or unexpected outcome or testing a development hypothesis underlying the program logical framework. Monitoring is the responsibility of the implementation team, while different stakeholders—including funding agencies, governments, and development professionals—share the evaluation challenge. The responsibility of the implementation team is limited in this case, although the availability of good implementation data is crucial for conducting a quality evaluation.

3.2 Types of Evaluations

There are different kinds of evaluations and also different typologies used to distinguish them. One such typology is whether the evaluation is internal or external: is the evaluation conducted internally by the implementation team (“self-evaluation”) or is it done externally by an independent party. Another broad distinction to consider is whether the evaluation is looking at operational issues or at the outcomes or net impact of the interventions. An evaluation that looks at operational issues (whether the interventions were implemented as planned, how resources were allocated, etc.) will typically review implementation documents, complemented or not by interviews with stakeholders. Rarely will it involve collecting additional data or performing complex empirical analysis. Evaluations that are concerned with realized or unrealized outcomes and the reasons behind those outcomes will usually require collecting data to compare the before and after situation. How systematic the data collection plan is and the extent of the analysis performed can vary greatly however even within that group of evaluations.

---

There is not a clear consensus in the literature on how to name the different kinds of evaluations. The terminology and definitions proposed by the United States Government are adopted here, as they serve the needs of this module well. The USG distinguishes between process, outcome, and impact evaluations as follows:

**Process (or Implementation) Evaluation:** This form of evaluation focuses on program or intervention implementation, including, but not limited to access to services, whether services reach the intended population, how services are delivered, client satisfaction and perceptions about needs and services, and management practices. It assesses the extent to which a program is operating as it was intended. It typically assesses a program activities’ conformance to statutory and regulatory requirements, program design, and professional standards or customer expectations. *Example of questions asked:* Are activities delivered as intended and are the right participants being reached?2

**Outcome Evaluation:** This form of evaluation assesses the extent to which a program achieves its outcome-oriented objectives. It focuses on outputs and outcomes (including unintended effects) to judge program effectiveness, but may also assess program processes to understand how outcomes are produced. An important element of Outcome Evaluation is that it measures short-term and intermediate term changes in the population being served. Furthermore, because study designs are sometimes taken from non-probability samples and samples of convenience, we cannot always conclude that a particular program lead to the result but did, in fact, contribute to any changes measured. Often in outcome evaluations there is no ability to control for confounding variables and the effects of other projects or programs on the target population. *Example of questions asked:* To what extent are desired changes occurring and who is benefiting?3

**Impact Evaluation:** Impact evaluations measure the change in outcomes that are long-term results of a project and they are designed in such a way as to be able to attribute measured changes in a target population to a defined intervention by comparing actual results to what would have happened in the absence of the intervention (the counterfactual scenario: see section 3.4 of this Module). Impact evaluations are based on models of cause and effect and require a rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. The ability to attribute any changes measured directly to one program/project is an important distinction between impact evaluations and outcome evaluations. Typically, comparisons are made between beneficiaries that are randomly assigned to either an intervention or a control group which provides the strongest evidence of a relationship between the intervention under study and the outcome measured to demonstrate impact.4 *Example questions asked:* Did this project reduce the rates of malaria among children less than 5 years of age? Is the average life expectancy of livestock in the intervention area increasing more than the average life expectancy of livestock in the comparison area?

**Economic Evaluation:** This type of evaluation relies upon the use of applied analytical techniques to identify, measure, value and compare the costs and outcomes of alternative

---

2 Ibid.
3 Ibid.
4 Ibid.
interventions. It is a systematic and transparent framework for assessing efficiency focusing on the economic costs and outcomes of alternative programs or interventions. This framework is based on a comparative analysis of both the costs (resources consumed) and outcomes (health, clinical, economic) of programs or interventions. Main types of economic evaluation are cost-minimization analysis (CMA), cost-effectiveness analysis (CEA), cost-benefit analysis (CBA) and cost-utility analysis (CUA). Example of question asked: What is the cost-effectiveness of this intervention in improving patient outcomes as compared to other treatment models?\footnote{Ibid.}

The cost model allows for evaluators to determine the cost of a particular benefit or effectiveness or utility of an intervention. These tools require that benefits, utility and effectiveness be monetarily quantified. While the accounting world has given us ways to quantify most costs, although not all, benefits, utility and effectiveness are not always easy to measure never mind cost out. The value of this analysis is that in the end, decision makers can compare different benefit/utility/effectiveness to cost ratios for different approaches and decide where the money is best spent; where do we get the largest benefit for the least cost. Which projects bring about the greatest utility, for what cost? Does effectiveness vary along with cost? Embedded in this is a clear understanding of decision-maker values and how much they are willing to spend on an outcome given its benefits, utility and effectiveness for the beneficiaries. There are a number of resources available on the Internet on economic evaluation analysis, although few apply to development projects.

3.3 Factors to consider when deciding on an evaluation

There are a number of factors one should consider when deciding on the type of evaluation. These factors include what questions are of interest or need to be answered, who the audience is (who is interested in the evaluation and who will be using the results and recommendations), and what resources in terms of funding, time, and expertise are available. Specific constraints, such as uncooperative stakeholders, problems with baseline data, or unexpected changes in the implementation plan, may also influence what kind of evaluation can be conducted.

The questions that an evaluation seeks to investigate should be the first matter to decide. It is a good idea to start deliberating about evaluation questions early, during project design, when the log frame is being developed and consultations with stakeholders are held. The process of identifying evaluation questions should include a broad representation of stakeholders to ensure that the evaluation addresses widespread concerns and interests and that it receives the necessary support while it is being conducted. Evaluation questions may need to be adjusted during implementation and not all questions of interest can be tackled by one evaluation, however it is useful to know the menu of evaluation questions relevant for a project, even if other constraints limit what can be effectively addressed.

Different evaluation methods can require vastly different amounts of resources and thus what resources are available for a particular evaluation should be clearly established. Trying to cut costs often means cutting corners and producing an evaluation that has
weak findings and ineffectual recommendations. Data collection is expensive, but a survey done with insufficient resources will most likely generate poor quality data that can end up being totally unusable. It is best to adopt an evaluation approach commensurate with the resources available.

It is also important to consider who the audience for the evaluation is. If the senior management requests an evaluation because added resources had to be channeled to a project during its implementation, the audience interested in the findings and recommendations of this evaluation is likely to be relatively small and a lot of outside expertise may not be necessary. If however, a new project offers an opportunity to design an impact evaluation with a strong counterfactual and an expert team is available, it can be expected to draw a large and sustained audience, as the findings may contribute to the body of evidence about aid effectiveness.

Below is a table summarizing the characteristics of the three main types of evaluations, along with various factors to consider when implementing an evaluation.
### The three types of evaluations and their characteristics

<table>
<thead>
<tr>
<th>Factors to consider</th>
<th>Process evaluations (Questions related to the activities being carried out by staff/volunteers)</th>
<th>Outcome evaluations (Questions focused on changes in the lives of the beneficiaries)</th>
<th>Impact evaluations (Questions related to the goal of the project and changes in community)</th>
</tr>
</thead>
</table>
| **Evaluation questions** | How closely did the actual implementation follow planned operations?  
Were the operational resources sufficient to implement the planned activities?  
To what extent were the objectives attained?  
To what extent were beneficiaries satisfied with the delivery of products and services? | What changes have occurred among the target population?  
- Knowledge  
- Skills  
- Behaviors  
Was the design of the project adequate to address the constraints identified? | What is the net impact of the interventions on the well-being of beneficiaries?  
To what extent are project beneficiaries better-off as a result of the project compared to non-beneficiaries  
To what extent can results observed at the end of the project be attributed to the interventions? |
| **Resources requirements** | Often relies on routine data that is already being collected.  
Costs are usually relatively low. If extensive qualitative research is undertaken with beneficiaries, time and expertise requirements will add to the tab. | Because it includes baseline and end line data collection, the survey costs and data analysis time and expertise need to be considered.  
And, if the evaluation includes qualitative research components, additional costs will be incurred. | The cost of an impact evaluation is typically higher than other kinds of evaluations because it requires rigorous data collection methods, more time to pass, and specific expertise for the design and data analysis.  
Cost magnitude depends on the complexity of the evaluation design and extent of data collection requirements, including possible qualitative research components. |
<table>
<thead>
<tr>
<th>Factors to consider</th>
<th>Process evaluations (Questions related to the activities being carried out by staff/volunteers)</th>
<th>Outcome evaluations (Questions focused on changes in the lives of the beneficiaries)</th>
<th>Impact evaluations (Questions related to the goal of the project and changes in community)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This kind of evaluation uses mainly administrative and monitoring data and requires limited data analysis. A desk study will use all project documents but will require limited data analysis. If the evaluation includes collecting information directly from beneficiaries, data analysis will be more extensive if sophisticated qualitative research methods are used.</td>
<td>Since it is an empirical analysis of outcomes, baseline and end line data collection and analysis are required. It may not require probability samples but should include some form of comparison over time and comparison to a non-exposed population if possible. A qualitative research element to the evaluation will have its own data collection methods and analysis requirements.</td>
<td>Impact evaluations always require relatively extensive data collection and analysis. How extensive depends to a large extend on the design of the evaluation, whether experimental or non-experimental methods are used, the size of the control or comparison group, etc. It must be rigorous enough to attribute observed changes to the program. A qualitative research element to the evaluation will have its own data collection methods and analysis requirements.</td>
</tr>
<tr>
<td></td>
<td>This kind of evaluation can be done relatively quickly and launched anytime during the project implementation or shortly upon completion.</td>
<td>Outcome evaluations require a great deal of time for planning and implementation. The length of time depends on the length of time that the project will run. Outcomes sometimes take a while to emerge from a population. Some outcomes, like shifts in knowledge, can be measured over a few months or a year. Other outcomes, like changes in farming techniques may require more time to be measured.</td>
<td>Evaluators will be in the field at various points in time during the implementation for data collection and staying abreast of implementation issues. The timing of the end line survey, which depends on the time it may take for the impact of the project to be realized, will determine the timeline for the completion of the evaluation.</td>
</tr>
<tr>
<td>Factors to consider</td>
<td>Process evaluations (Questions related to the activities being carried out by staff/volunteers)</td>
<td>Outcome evaluations (Questions focused on changes in the lives of the beneficiaries)</td>
<td>Impact evaluations (Questions related to the goal of the project and changes in community)</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Audience</td>
<td>This kind of evaluation is most useful to implementers and funding agencies as it helps with decisions regarding resource allocation and strategic orientation. Usually, the findings and recommendations of these kinds of evaluations are of limited interest to a wider audience.</td>
<td>This kind of evaluation can have a broad audience beyond the immediate stakeholders. It can provide useful information on the underpinnings of project logic and on measuring and analyzing project outcomes.</td>
<td>This kind of evaluation usually has a broad and sustained audience. It has the potential to add to the body of evidence on what works and what does not work and suggest new areas of interventions and research.</td>
</tr>
</tbody>
</table>

The following figure demonstrates that not all programs are involved in all levels of M&E. While most or all programs will conduct various aspects of the Process M&E as we move along into Outcome and Impact evaluations fewer and fewer programs will be directly responsible. This figure also attempts to further delineate the differences between each level of evaluations.
Figure 1-1: Monitoring & Evaluation Pipeline

**MONITORING**
"Process Evaluation"

- Inputs
  - ALL
  - MONITORING
    - Resources
    - Staff
    - Funds
    - Materials
    - Facilities
    - Supplies
    - Training
  - Outputs
    - Condom availability
    - Quality of services (e.g., STI, VCT, care)
    - Knowledge of HIV transmission

**EVALUATION**
"Effectiveness Evaluation"

- Outputs
  - Most
    - Short-term and intermediate effects
      - Behavior change
      - Attitude change
      - Changes in STI trends
      - Increase in social support
  - Some
    - Long-term effects
      - Changes in:
        - HIV/AIDS trends
        - AIDS-related mortality
        - Social norms
        - Coping capacity in community
        - Economic impact
  - Few

**Levels of Evaluation Efforts**

**Number of Projects**
- ALL
- Most
- Some
- Few
3.4 The counterfactual: the key to attribution

The counterfactual in evaluation is the situation that would have prevailed had the project not taken place. What would have been the incomes of the beneficiaries, for instance, without the interventions? This is of interest because ultimately, we want to know that the outcomes observed at the end of a project are indeed due to the project and that they would not have been achieved without the interventions. Because projects are not implemented in a vacuum or in a laboratory where external factors can be controlled, there are constantly events that can influence positively or negatively the behavior of the people and institutions we work with. We may be aware of some of these events, such as the presence of another project or the launch of a government program. There are other events, however, that we may not be aware of or that may not be observable. There may be some inherent characteristics of the participating households or farmers that we are not aware of, such as risk taking ability or self-confidence, or patterns of information sharing among communities that are not apparent. Even if we are aware that behavioral patterns or unpredicted events have had an influence, we do not know the size of their effects on the outcomes.

The “without project” situation cannot be reproduced once a project is implemented. However, statistical methods can be used to estimate the counterfactual through the use of control or comparison groups. The ability to identify a statistically acceptable control or comparison group is at the core of an impact evaluation design. Note that when a strict distinction is made between control and comparison groups, it is usually meant that a control group is selected randomly from the same population as the program participants, while a comparison group is not random (or not entirely), but can be statistically compared to the participating population. This distinction is also used to characterize experimental and non-experimental methods. An experimental design will use a control group identified through a randomization process (randomized design), while a non-experimental design will use other methods to identify a comparison group. The statistical procedures that are used and the conclusions that can be drawn from a randomized design are different from what can be derived from a non-randomized design, and the former is considered as the more rigorous approach. More in-depth discussions on experimental and non-experimental methods used for impact evaluations can be found in some of the references section.

The figure below illustrates the difference a valid counterfactual has when measuring the impact of a project. Without the counterfactual, the impact (income in this case) is erroneously measured as the distance between points B and A, that is, the level of income of the beneficiary population at the end of the project minus the baseline value. However, even without the project, incomes would have increased. Through the identification of a control group which does not participate in the project but is statistically similar to the population participating, the income level without the project can be estimated (point C), and the true net impact of the project can be estimated to be the distance between points B and C.
3.5 Putting together a protocol and SOW for an evaluation

Once a decision has been made on the type of evaluation that can best investigate the questions of interest with available resources, a protocol needs to be developed and eventually turned into a scope of work. The protocol, which presents the evaluation questions to be addressed (project logical framework) and the research methods to be conducted (survey, focus group, key informant interviews, see Module 8), is the first step in ensuring a quality evaluation that will be useful to the project. At a minimum protocol includes:

- Background/situation in the target community
- Program/project description and objectives
- Evaluation purpose and objectives
- Key questions and indicators
- Methodology; including sampling and research approach
- Consideration of ethical reviews
- Logistics
- Budget
- Dissemination and use

A template for a protocol can be found in Annex 9.8.

The best way to ensure that the SOW includes all relevant items is to base it on the protocol for the evaluation. Often it is easiest to attach the protocol to the contractual document that establishes the arrangement between Africare and the research entity.
If the funder is USAID then the implementing organization may not be tasked with drafting the SOW for an evaluation. This depends on the level of the evaluation and the size of the program budget in relation to other USAID program budgets. USAID has retained the responsibility for all impact evaluation solicitations and depending on the size of the budget many of the outcome evaluations, including the drafting of SOWs. However, implementing organizations are pivotal partners in any evaluation and should be fully engaged in the process to ensure that evaluations will be useful, are conducted with a clear purpose and in a collaborative and transparent manner, collecting quality data, and able to reach useful recommendations. And if the project is funded through a mechanism other than USAID, Africare may have full responsibility for all levels of evaluations.

There are a number of references available on the Internet that can help with planning and drafting a SOW for an evaluation but your protocol is always the best place to start. USAID has produced a series of documents following the release of its Evaluation Policy (January 2011). These documents are useful tools for missions and USAID partners in planning and conducting evaluations and a number of them are listed in the Reference and Recommended Reading section below.

As an example of how to put together or how to read an existing SOW, we reproduce below in the form of a checklist, the USAID requirements as stated in the ADS Chapter 203.

**Checklist**

USAID ADS 203.3.1.5 states that a well-written statement of work should:

- Describe the specific intervention, project/program, or process to be evaluated
- Provide a brief background on the development hypothesis and its implementation
- Identify existing performance information sources, with special attention to monitoring data
- State the purpose of, audience for, and anticipated use(s) of the evaluation
- Identify a small number of evaluation questions that are relevant to future decisions and answerable with empirical evidence
- Identify all evaluation questions for which gender-disaggregated data are expected; also identify questions for which an examination of gender specific or gender differential effects are expected
- Calculate sample sizes by segmented target populations and explain how sample frames will be created if they do not already exist
- Identify evaluation method(s) that will generate the highest quality and most credible evidence on each evaluation question, taking into account time, budget, and other practical considerations and specify methods with sufficient detail
- Describe how data collected on evaluation questions will be analyzed
- Describe strengths and limitations of the evaluation methods
- Explain how Informed Voluntary Consent will be obtained from potential respondents and identify who is responsible for obtaining the Ethical Review approval, Africare or research partner
Specify the evaluation deliverable(s) and their timelines and logistics, including requirements for the transfer of data to USAID and expectations concerning evaluation team involvement in the dissemination of evaluation results.

Clarify expectations about the methodological and subject matter expertise and composition of the evaluation team, including expectations concerning the involvement of local evaluation team members (one team member should be an evaluation specialist).

Describe intended participation of USAID staff, implementing partners, national counterparts or customer/beneficiaries in the design or conduct of the evaluation.

Address scheduling, logistics and other support.

Clarify requirements for reporting and dissemination.

Include a budget.

REFERENCES AND RECOMMENDED READING


USAID Resources:

- **USAID Evaluation Policy**, January 2011:
MODULE 4  PROJECT INDICATORS

PURPOSE OF THE MODULE

Indicators are deceptively complicated. They need to be very concise and precise and yet tell an entire story to whoever collects them. If the project has been designed with SMART result statements and objectives then the selection or design of the indicators can be very easy and straight forward. Often it is at the point that we start choosing indicators that project results and objectives take on an improved format. Indicators must always be related directly back to the results and objectives of a project. Therefore, the best place to start is with the Log Frame or results framework. The Log Frame shows the hierarchy of expected outputs and outcomes of a project that should contribute to a desired impact. From this framework, a set of indicators can be defined and targets established. Indicators are the measure and standards by which each result is to be assessed. There are different types of indicators.

CONTENT OF THE MODULE

4.1 Selecting, Testing, and Using Indicators 25
4.2 Custom vs. Standard or Common Indicators 25
4.3 Indicator Level 26
4.4 Characteristics of Good Performance Indicators 26
4.1 Selecting, Testing, and Using Indicators

Indicators are data points that are used as clues, signs, and markers to show us how close we are to our path and how much things are changing. They point to, or indicate, possible changes in the situation that may lead to an improved result. For example, if you are driving in a car and the gas gauge shows you are low on gas, it is not actually the gas you are looking at, but rather you are looking at an “indicator” of the amount of gas you have. And with that information you can now make an informed decision on what to do next, data for decision making.

Therefore, the selection of your indicators is crucial to the ability to manage and learn from your project’s data system. Pick the wrong indicators and you will miss the opportunities to understand what is going on.

As you begin the process of choosing and designing indicators, each indicator should be assessed using the following questions:

- Have the definitions of the indicators been tested so that it is doable, operational?
- Can it be measured accurately each time it is measured, reliability?
- Will the indicators measure only what they are supposed to measure, validity?
- Are there areas of overlap in the content of the indicator with that of other indicators, or is it too general, specific?
- Will the indicators be able to measure changes over time, sensitivity?
- What resources (human and financial) do the indicators require, affordable and feasible?
- Are there alternative measures that should be considered, proxy measures?
- Will multiple indicators be able to help clarify the results of the primary objective?

Indicators fall into several categories:

- Project customized
- Internationally standardized
- Level of result or process (Input, Output, Outcome, Impact)

4.2 Custom vs. Standard or Common Indicators

Custom indicators are those that are specific to a project and capture progress of particular activities. Their purpose is to capture changes at the level and pace that matter locally. Standard or common indicators on the other hand are indicators that can be aggregated across projects or countries. They are used mostly for agency reporting to demonstrate the overall reach and scope of the programs implemented or supported. Such indicators are usually defined independently of any particular project. They attempt to capture commonality rather than specificity because they can be aggregated across many projects, defined in a standard way, and defined in units that can be added (such as number of individual, dollar spent, or km of roads). While these two types of indicators appear to have different uses they are both crucial to good solid management; standardized indicators would not have been selected as good indicators if they did not prove to be useful for program design and management.
4.3 Indicator Level

A common source of confusion in indicator selection is the “level”, or whether the indicator is measuring an input, output, outcome, or impact results. This is important as we are working in a hierarchical system and its coherence depends on the lower level indicators feeding information into the next higher level. Only the highest level can be an end in itself. An indicator that fails the level test and is unclear as to what its level is or what information it feeds into the next higher level should be discarded.

Input indicators should measure work plan milestones, resources and supporting activities that take place prior to implementation. Output indicators directly measure the activities carried out by project staff. Outcome indicators measure results at the beneficiary level, while impact indicators measure the change in well-being that was sought by the project and can be considered an end in itself in our hierarchical system.

### EXAMPLES of appropriate indicator level structure

<table>
<thead>
<tr>
<th>Input:</th>
<th>5,000 bed nets have been purchased (Date)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output:</td>
<td>Number of bed nets distributed to households with school-aged children</td>
</tr>
<tr>
<td>Outcome:</td>
<td>Percentage of school-aged children that sleep under a bed net every night</td>
</tr>
<tr>
<td>Impact:</td>
<td>Percent reduction in reported malaria cases in school-aged children</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input:</th>
<th>Trainers have been trained in selected improved rice cultivation techniques (Date)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output:</td>
<td>Number of farmers trained in improved rice cultivation techniques</td>
</tr>
<tr>
<td>Outcome:</td>
<td>Number of farmers using improved rice cultivation techniques in at least half of their plots</td>
</tr>
<tr>
<td>Impact:</td>
<td>Percent increase in farmers’ marketable surplus of rice</td>
</tr>
</tbody>
</table>

4.4 Characteristics of good performance indicators

There are different models to establish what makes a good indicator, but all use similar criteria. The model used by USAID (ADS Chapter 203), provided below, uses seven criteria to guide indicator selection. Below is a table developed by USAID on how to use those criteria.

Another model is the CREAM (Clear, Relevant, Economic, Adequate, Monitorable) model (see in particular the World Bank, 2012 and ADB, 2007 in the Reference and Recommended Readings section).
No matter what model you use to choose your indicator, make sure you keep in mind what it will mean to track it. You will be doing the work, make it easy on yourself. An indicator should provide the information that you need to monitor your project, know how well your project is being carried out, meet your reporting requirements to management and the donor, and should not break your budget or your sanity. It should be clear enough to everyone what it means and why it is useful. If an indicator does not pass your test and that of the implementing team, drop it.

A common mistake teams make at the beginning of a project, when the detailed planning is under way, is to want to collect data on everything and monitor every possible indicator. It is important to be disciplined in selecting indicators, as tracking them is a lot of work. Too many indicators are a burden and will at the end jeopardize the usefulness of the system.
### INDICATOR SELECTION CRITERIA TABLE

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Direct</td>
<td>The indicator clearly represents the intended result. An outsider or an expert in the field would agree that the indicator is a logical measure for the stated result.</td>
</tr>
<tr>
<td></td>
<td>o Level: The indicator reflects the right level; that is, it does not measure a higher or lower level than the stated result.</td>
</tr>
<tr>
<td></td>
<td>o Proxies: The indicator is a proxy measure. If the indicator is a proxy, note what assumptions the proxy is based upon.</td>
</tr>
<tr>
<td>2. Objective</td>
<td>The indicator is clear and unambiguous about what is being measured.</td>
</tr>
<tr>
<td>3. Useful for Management</td>
<td>The indicator is useful for management decision-making.</td>
</tr>
<tr>
<td>4. Attributable</td>
<td>The indicator can be plausibly associated with USAID interventions.</td>
</tr>
<tr>
<td>5. Practical</td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Time: Data are produced with enough frequency for management purposes (i.e. timely enough to correspond to USAID performance management and reporting purposes). Data are current when available.</td>
</tr>
<tr>
<td></td>
<td>o Cost: Data are worth the cost to USAID managers.</td>
</tr>
<tr>
<td>6. Adequate</td>
<td>The indicators, taken as a group, are sufficient to measure the stated result. All major aspects of the result are measured.</td>
</tr>
<tr>
<td>7. Disaggregated, as necessary</td>
<td>The indicators are appropriately disaggregated by gender, age, location, or other dimensions considered important for programming. In particular, gender disaggregation has been considered as required (see ADS 203.3-4.3).</td>
</tr>
</tbody>
</table>

*Source:* USAID TIPS #6, 2010, p. 11.

---

6 For Africare, gender and age are the most important disaggregation.
MODULE 5 USE OF THE RESULTING M&E INFORMATION

PURPOSE OF THE MODULE

Up to this point we have reviewed definitions, outlined systems, looked at different levels of M&E, and determined what to measure. In this module, we will cover the various uses of the data that results from your M&E system.

This is important to be thinking about from the very conception of a project, when an RFA/P comes out it is important to start thinking about what needs to be told about this project and who will need to hear it. The better understanding we have, at this stage, of the uses for this data the better understanding we have of the resources that will be needed (human and financial) when the project is up and running. If there is an outcome level evaluation needed then you will need particular skills and sufficient funds to carry out a study. If not, then there are other needs to be met. If this is an innovative approach we will want to show its utility and be able to replicate it in other areas and that will require certain types of evaluations be carried out.

In the end, early planning and thinking will ensure that each project has the right system designed, in place and financed to meet the ultimate needs of telling our story as best we can.

CONTENT OF THE MODULE

5.1 PROJECT SELF-ASSESSMENT
5.1.1 Process Evaluation
5.1.2 Outcome Evaluation
5.1.3 Quality assessments
5.2 PROJECT EXPANSION AND REPLICATION
5.3 MARKETING OUR SUCCESSES
5.4 REPORTING
5.1 Project self-assessment

The first use of this data should be for the project staff to look at themselves and see where we are within the implementation of the project. To accomplish this we need to ensure that the right data is being collected that provides a good overview of what we are doing. And then we need to make sure that routine meetings are scheduled to present the data and offer an time for discussion and questions. These meetings should be held at least every month but if staff is widely scattered and transportation is difficult they could be held each quarter.

In these meetings we want to be able to show our progress to date for each indicator that we are following and compare it to the targets that we have agreed to with the funder. Where targets are being met, or progress towards targets is on track, we have an opportunity to talk about what is going well. At the same time, when progress is not on track we have an opportunity to discuss what is happening that is impeding our progress. The earlier we have these discussions (hence the monthly meeting) the sooner we can intervene on issues that are blocking our successes and make adjustments to the project.

5.1.1 Process Evaluation

This first level of data use falls into what can be called the “Process Evaluation”. In this type of evaluation we are looking at stated targets for each objective and comparing those targets with measurements taken on those the indicators that have been established to monitor those objectives. For example, for an agriculture project we have the following bits of information:

| Objective: | To improve veterinary care |
| Indicator: | # of veterinary interventions (treatments and vaccinations) |
| Q1 | Q2 | Q3 | Q4 | Yr. 1 | Life of Project |
| Targets: | 0 | 22,982 | 35,000 | 4,136 | 62,118 |
| Actual results: | 0 | 16,950 | 38,654 | 474 | 56,078 |
| % of Q or Yr | 74% | 110% | 11% | 90% |

To be clear, this is not the only indicator to measure improvement in veterinary care, but it is one that at least tells project implementers something about the reach of the project and how closely the project is coming to achieving targets agreed upon with the funder. The decision as to whether achieving 90% of an annual target is good or not is a decision that each project must make in collaboration with their funder. It would be a very good idea at this point to be able to explain any anomalies in the data. For example, the fourth quarter barely made it past the 10% mark for the number of veterinary interventions conducted. What happened during the fourth quarter? How do we explain this result?

This first level of data use does not require sophisticated data analysis packages or extensive data collection systems. Project staff keep track of the interventions being carried out daily and aggregate that on a weekly, monthly and finally quarterly basis either on paper or in an electronic database and then make the comparison to the targets. On the other hand, if we are interested in knowing what the effect of these veterinary visits
have on the livestock we will need to look at a higher level of indicator and a more complicated data collection system.

### 5.1.2 Outcome Evaluation

At the outcome level, recall from previous modules, we are interested in finding out what the effect on our target population might be as a result of participation in our project. This requires a more rigorous approach to data collection and therefore requires more human and financial resources. Let’s look at the same livestock project...

<table>
<thead>
<tr>
<th>Objective:</th>
<th>To improve veterinary care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator:</td>
<td>% of beneficiary households who have seen a stabilization or an increase in the size of their livestock thanks to the support provided by the project</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Yr. 1</th>
<th>Life of Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>60% of HHs</td>
</tr>
</tbody>
</table>

This is not routine data and it can take some time for this effect to show up in the actual herds of the beneficiaries; it is affected by number of live births, survival rates and adult mortality in the herd. Therefore, a more complicated system to collect baseline and end line, midline for longer projects, needs to be developed and resources allocated.

By bringing together the process indicators (i.e., # of veterinary interventions, # of animals benefiting, # of animals treated, quantity of animal feed provided, # of animals restocked etc...) and outcome indicators such as % of households with stabilized or increased herd sizes we will be able to know if our intervention is working and can make decisions about either changing/improving the design or expanding the designed project to new areas in need.

### 5.1.3 Quality assessments

Another very important part of the self-assessment is an assessment of the quality of the work that is being carried out. It doesn’t matter how many cows are treated by the veterinarian if the veterinarian is not performing up to standard, or if the medication that is being used has already expired, or if the material needed to perform the exams is broken. Above and beyond the actions being carried out each project needs to have the ability to observe and measure the quality of the interactions with beneficiaries. Is our staff, or are our volunteers, carrying out the program as design, with skill and proper equipment?

### 5.2 Project expansion and replication

Once we have described our project’s process and outcomes we are in a position to take effective projects to new areas in need. The current funder may be interested in adding in new villages or districts and with the success of our project we might be awarded new funds to expand into surrounding areas. At the same time a well-designed and
documented project can be used to support a brand new proposal in a different area altogether. It is important to keep in mind, however, that cultural differences dictate moving slowly and consciously when expanding to new cultures and communities. What works in Nigeria may not work in Tanzania, however, the basic idea might be transportable with cultural adaptations built in.

For example, Africare conducted a highly successful HIV prevention project with Female Sex Workers in Nigeria. The basic ideas of this project were brought into the discussion of a response to an RFA by a funder looking to start another project in Tanzania working with young women engaged in high risk sex through exchange of money for sex. The Nigeria project was designed to enable vulnerable women to live healthier lives through training on Reproductive Health/HIV/AIDS and treatment of STIs and to economically empower women through vocational and business management training and the provision of start-up equipment. Because clear and detailed summaries of the project and outcomes were available, this project was helpful in designing another project to be carried out with the same type of population but in another geographic location.

5.3 Marketing our successes

Marketing is not often the first thing that people think of when thinking about their M&E system. And yet, if we do not do a good job telling the world who we are and what we do no one else will. Our data can be a rich and valuable source of information to share with potential donors and potential partners. We need to be able to show the world who we are, what we do and how well we can do it.

As projects develop their M&E systems one thing to keep in mind is how we will be able to combine similar projects across countries and regions for marketing and reporting purposes. If we are conducting nutritional programs for children under 5 then we have an opportunity to show greater levels of “reach” if we have written the objectives in similar ways and have indicators that lead to that objective, even if not the same indicators. The related indicators for each project can then be collapsed across projects.

For example...

<table>
<thead>
<tr>
<th>Objective: Improved nutritional status for children &lt;5 years of age</th>
<th># of Children reached in 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td># of children &lt; 6 months exclusively breastfeeding</td>
</tr>
<tr>
<td>Zambia</td>
<td># of children &lt; 5 eating vitamin A enriched mielie meal</td>
</tr>
<tr>
<td>Uganda</td>
<td># of 6 – 23 months old with access to high-quality diverse foods</td>
</tr>
<tr>
<td>Nigeria</td>
<td># of children &lt;5 years old with access to clean drinking water</td>
</tr>
<tr>
<td><strong>Total Children reached</strong></td>
<td><strong>47,351</strong></td>
</tr>
</tbody>
</table>

Even though the indicators are not the same they all feed up into the objective of nutritional support and now Africare can report that we reached 47,351 children under the age of 5 with interventions and activities designed to support improved nutrition. If any of
these projects did further research on actual changes in health status then we can supplement these overall numbers with real improvements. But even having this reach number is useful because it shows a strong emphasis on a particular issue and can also be supplemented with a few individual stories or case studies.

5.4 Reporting

Writing a report to send to the funder tends to be the first thing people think of when they think about the M&E system, so it has been saved for last. But that does not mean it is not crucial to our work. The funder requires reports on varying schedules so it is important to know what each funder is asking for: monthly, quarterly, annual, end of project...

The data within your M&E system should be thought of as a way to tell the story of what your project is accomplishing, or has accomplished, with the funders money. They have a right to know what we are doing and by turning in quality timely reports we can ensure that the funder is happy and we increase the possibility of getting additional funds to support our good work.

It is important to be aware of not just the periodicity of the required reports but also be aware of the requested information. During the first several months of a new project the funder and Africare will come up with an agreed upon list of indicators. This list, and progress to date, must be a part of any report to the funder. Along with the progress being made on each indicator it will be important to be ready to explain any extremely low or high rates of progress. For example, if we are half way through a project and we report to the funder that we have reached 20% of a target then we need to be able to explain what the barriers have been to achievement. Likewise, if we are halfway through a project and we have already achieved 100% of the target then we need to be prepared to explain what happened to get us there so quickly: ensure that the data is correct, did we make a target error, or did something out of our control occur to increase interaction with our project?

And of course this same set of data can be used to report to HQ on progress being made an to the Africare Board of Directors so that they get a good solid understanding of what is happening and who we are reaching.

First and foremost think about this data as an opportunity to tell the story of Africare’s experience with this project and in the community.
MODULE 6      THE PRE-AWARD PHASE

PURPOSE OF THE MODULE

This module describes the steps that are required during the proposal development phase before an award is granted. Some of these steps are strictly confined to that stage, while others may be started at that stage and completed in the next stage, once an award is obtained and implementation begins. M&E should play an active role during the proposal development process, because this is when decisions are made about the objectives of the project and logic of the interventions, which essentially set the performance parameters (and how easy it will be to monitor them). This stage is critical to ensure that M&E is effectively integrated in the project cycle and that it remains relevant and useful throughout implementation.

There is an often lengthy time lag between project design and the start of implementation, during which time crucial information can easily be lost or forgotten. It is therefore important to carefully document what is being done during the pre-award phase, how and why decisions are made, and what follow-up steps are recommended to maintain overall coherence.

Africare follows a “Problem Analysis Framework” approach to design a new project and develop a proposal. This approach is described in the “Proposal Development Toolkit” (September 2011). The framework offers a step-by-step process to analyze a problem in a logical system, define the main objectives, and identify the components of an intervention strategy. We draw attention in this module to a number of trigger points for M&E.

CONTENT OF THE MODULE

| 6.1 | ARTICULATING A RESULTS STATEMENT | 35 |
| 6.2 | CONSTRUCTING THE PROJECT LOGICAL FRAMEWORK | 36 |
| 6.3 | PRE-AWARD M&E PLAN/PROPOSAL | 39 |
| 6.4 | BUDGETING FOR M&E | 40 |
6.1 Articulating a results statement

During the pre-award phase, the development issue put forward by the solicitation request is analyzed in terms of what the problem is, what are its likely causes and effects, and what are ways to mitigate or solve the problem. The process leads to articulating a series of results statements from the overall goal to specific inputs, outputs, and outcomes identified as intermediate steps to reach the goal.

A results statement should be clearly articulated around one objective—not a series of objectives or expected results—and be measurable.

**EXAMPLES of good and poor results statements**

**Increased crop yield of rural households in the targeted areas**

**Reduced incidence of malaria of children under 5 years old in the targeted areas**

Both of these goals are acceptable results statements. They are expressing direct, single objectives that are achievable and measurable.

**Improved well-being of households in the targeted areas**

**Increased food security and improved nutritional status of households in the targeted areas**

The statements above are not well-formulated results statements. In the first case, household well-being is a vague objective that cannot easily be measured (unless we mean income, in which case, it should be in the statement). Although it may be attractive because it appears to encompass a broader goal and the project may aspire to touch households beyond an income gain, it is important to remain practical when articulating what a project wants to achieve. It is better to have an objective that can be defined and measured without ambiguity.

In the second example, there are two outcomes in one statement, giving rise to ambiguity as to what is more important and how success will be characterized if one objective is met and the other is not. One of the two objectives should be chosen as the priority one.

An overall effective way to ensure that your statements are useful is to follow the SMART approach to results and objective writing. In this approach we write out statements that are:

- Specific
- Measurable
- Attributable
- Realistic
- Time bound
A statement or objective is specific if it is focused on just one result or activity. Sometimes, as in the second example above, we clump more than one thing together (food security and improved nutritional status, for example). By mixing them both into one result we cannot measure which, if either has been achieved. It is better to separate them out and have a result/objective for each one.

Measureable means that we can actually set up systems, with resources and tools, to determine that something is taking place. This means that it must be something that can be measured. For example, until recently measuring incidence of HIV in a community was too resource intense so it could not be measured. Likewise, improved well-being is too vague and therefore no direct measure of that is possible. Often the result statement defines the measurement itself by providing the end line target: By the end of the project there will be a 55% increase in the number of cattle being annual vaccinated in the project region.

Attributable looks at the degree to which non-project factors may be influencing and affecting the variables of outcome within the beneficiary population. The big issue here is whether or not there are other similar projects targeting the same beneficiary population. But also questions whether or not a result is something that a program can or will be able to effect. Is there a logical and real relationship between actions of the program and the intended result? If we are exclusively working to increase animal vaccinations in a pastoral community can we claim that this will result in increased nutrition among < 5 year olds? It may be a contributor but it is not the change that will help.

Realistic is also related to the idea that our actions are linked with the stated result or objective. If we are teaching communities to dig and care for wells, can we claim that that is what caused a decrease in cholera? Not because there are a number of issues that must be addressed to halt the spread of cholera through a community (clean sanitation sites and hand washing, for example).

Time bound means that the result statement clearly defines by when the result can or should be realized. Simple statements like, “By the end of the project...” or “By the end of year 2...” alert everyone to when we can expect to see these results manifest.

By making the result statements and written objectives SMART we have already gone a long way in identifying what will be monitored and evaluated. This is used to create the project log frame and then turns into the first part of your M&E plan.

6.2 Constructing the project Logical Framework

The entire team assigned to the design and proposal development will put together the project logical framework also called “Log Frame”, “results framework”, or “theory of change”. This step is pivotal for how M&E will be carried out during implementation. The project logical framework should:

1. Have a strong coherence,
2. Be based on available evidence or clearly spelled-out assumptions, and
3. Have a hierarchy of expected results that is unambiguous.
Everything should be documented. Although the project Logical Framework is a graphical representation of a proposed strategy, there should be a good bit of analysis and inferences made from the implementation of past projects (either by Africare or other implementers), and literature reviews about evaluation findings and other empirical research that go into this diagram. The project Logical Framework is the first step in developing a monitoring and evaluation system and documenting how it is constructed should be given proper attention.

Some important considerations in the construction of project logic are the following:

- Keep it simple. Too many layers and arrows are confusing. The purpose here is to simplify the argument. Not everything needs to be captured on the diagram and only the leading propositions should be included.

- If the project is large and relatively complex, it may be useful to develop more than one Log Frame. In these cases, some intermediate activities can be broken into stand-alone projects, allowing for clear representation through the use of smaller, individual Log Frames. This segmentation then allows each intermediate step to be linked together and clearly depict the process of the entire project, capturing the primary argument.

- Keep it one-directional. Feedback loops may be part of the development paradigm, but not of a results framework, which is a hierarchical model. The project Logical Framework describes how investing resources into a set of activities will lead to a series of specific targeted changes and eventually to an improvement in the well-being of a population. See Below.

- Review the evidence and document the assumptions. There is an expanding body of studies, evaluations, and other empirical research findings that are providing new understanding and innovative avenues about cause-and-effect relationships. Tools such as cost-benefit and beneficiary analysis can provide a solid empirical framework while putting together a project Logical Framework and help in identifying weaker assumptions or components where evidence is lacking. Some databases are being developed and are available on the Internet to facilitate search of what exists in terms of impact evaluations and other empirical analysis of development data.\(^7\)

\(^7\) A compendium of databases on impact evaluations that are available on the Internet was recently created by the Center for Global Development (CGDev): [http://www.cgdev.org/page/list-impact-evaluation-databases](http://www.cgdev.org/page/list-impact-evaluation-databases)
EXAMPLES of project logics

The first example below shows a project logic where the results statements are well-formulated and the hierarchical logic is clear.

The following examples are of what to avoid when putting together a results framework.

What is depicted above is not a hierarchical cause-and-effect relationship because the lower level results are not sub-categories of the higher level. The lower level results are in effect on the same level.

---

8 These examples are adapted from USAID TIPS #13. See References and Recommended Readings below.
Another frequent mistake is when the lower level is merely a restatement of the higher level (or vice-versa). This leads to more problems down the road because, in a situation like this, it is difficult to define indicators that are at different levels.

Another example with several concerns is shown below:

In the above example, the first lower level result is just a restatement of the higher level result. The second lower level result is not clearly in direct line with the higher level result and some explanation is needed to show the relationship. While the third lower level result is vague, unclear and needs a definition or what is meant by “livelihood systems”.

6.3 Pre-Award M&E plan/proposal

During the pre-award phase most funders are looking for a short description of what the M&E system will look like once awarded and an illustrative list of indicators. These need to be concise and precise so that they fit within the page limits of the funder’s requirements for the RFP and be able to clearly show that we have the resources, knowledge, and skill to conduct the necessary work.

The narrative of the M&E section of the pre-award proposal should always follow the
guidance set up in the RFP. If that guidance is not clear then the following is a good
generic guidance that can be incorporated into any funding mechanisms request. Be sure
and include a brief description of how we will program and monitoring interventions,
measure results, ensure data quality and use data to inform on-going implementation
and/or changes in direction.

The plan should highlight:

1) Considerations for baseline, mid-term, and/or final data collection and analysis,
2) Critical outputs, anticipated outcomes, and performance targets, indicator and
   benchmarks, and
3) Potential areas for program research and learning

Provided that there is sufficient space allotted to the proposed M&E plan, it is useful to
include an Illustrative Table of Indicators. (See the appendix for an example of a generic
table that can be used as is or adapted as needed.) This table is illustrative and will
become a piece of the full Performance Monitoring Plan once the award has been made. It
includes the list of indicators by project objectives and then provides the “source” of the
data, presumed baseline levels, annual targets, and life of project targets. In some
situations it is admissible to only provide end of project targets without the annual targets.
If a baseline survey is to be conducted then the baseline values for each indicator can be
labeled as “To Be Determined” (TBD). If, on the other hand, this is an extension of a
current project then you can site the final numbers for the previous project as the baseline
for the new project.

6.4 Budgeting for M&E

Setting a budget for M&E is challenging for several reasons. But the main reason is that we
want to make sure that the bulk of the money being spent on a project goes to
implementation rather than activities that support or surround implementation. However,
it is important to remember that the purpose of conducting M&E is to ensure that we are
implementing the best project possible and that our work is supporting the greatest
amount of positive change. Without these data a project manager really cannot honestly
tell if they are meeting objectives, targeting the right people, implementing to the
satisfaction of the target population, or even making a difference in people’s lives.

The USG has stated that they believe that between 5% and 10% of a project’s budget
should go to M&E. This is a good starting point but we have to be realistic. If the budget is
only $100,000/year then we are limited to $5 – 10,000 for M&E. That sets a limit on what
we can do, and rightfully so, since it is a relatively small budget. But even within this
limited budget there are things that can be done. Perhaps we only use $5,000 to set up and
manage the routine tracking data. That leaves an additional $5,000/year to carry out an
assessment at the end of the project. If the project is a 3 year project then we have $15-
30,000 to hire a consultant to carry out a small qualitative assessment to gain information
from partners and beneficiaries.

On the other hand, if we have a $10 million project you would probably not need $500,000
to $1 million to carry out M&E over the entire life of project. But for most of our projects
staying within that 5-10% range will allow you to hire staff and carry out both routine data
collection as well as special studies.

It is during this pre-award phase that you will decide how much of the budget will go to M&E. That means that you need to pay close attention to the indicators we are promising to report on and ultimately the methods that we will need to carry out in order to collection those indicators; these can really drive up the cost. If we tell a funder that “by the end of the project we will increase the food diversity eaten by <5 year olds in our target areas” then we have to have baseline and end line information; that means a survey and that costs money.

If a budget is over $1,000,000/year or more than $2 million for the life of the project then you required to have a full-time M&E person on staff. Any budget under that cut off can get by with a part of a manager’s time dedicated to routine data collection and data entry. One way to address this staffing issue for smaller budgets is to hire an M&E person who works on 2 or 3 smaller projects and has their time fully covered by those 2 or 3 projects.

One last budget item to build into the planning is the support from HQ. If your project requires any sort of special studies, baseline/end line, or additional assessments, whether they are quantitative or qualitative then you need to make sure that there is time for an HQ M&E backstop to not only provide support from Washington, DC but also to travel to country at least once each year to monitor that process.

REFERENCES AND RECOMMENDED READINGS

- http://betterevaluation.org/evaluation-options/CostBenefitAnalysis

Note: We have found that the direct link to the USAID TIPS documents did not always work. If you encounter a problem, the following website has links to all of them: http://www.innonet.org/resources/node/636
MODULE 7  THE PERFORMANCE MONITORING PLAN OR M&E PLAN

PURPOSE OF THE MODULE

The purpose of this module is to discuss the components of a comprehensive performance monitoring plan (PMP) or M&E plan that goes beyond a set of indicator tables. The PMP should be considered one of the strategic documents for a project, outlining the evidence and hypotheses supporting the results framework, discussing the assumptions and risks to achieving the expected results and how performance will be monitored. PMPs are normally drafted during the first 60 to 90 days of a new project and should be periodically updated as needed to ensure plans, schedules and assignments remain current.

When putting together the PMP, M&E Plan, write it keeping in mind someone at the funder level who wants to know exactly what is happening and when. Think about a new M&E person or manager coming on board 2 years into a long-term project; s/he can use this document to know exactly what is happening with M&E and when it needs to happen.

CONTENT OF THE MODULE

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>INTRODUCTION TO THE PERFORMANCE MANAGEMENT PLAN OR M&amp;E PLAN</td>
<td>43</td>
</tr>
<tr>
<td>7.2</td>
<td>COMPONENTS OF A PMP OR M&amp;E PLAN</td>
<td>43</td>
</tr>
<tr>
<td>7.3</td>
<td>PERFORMANCE INDICATOR REFERENCE SHEET</td>
<td>44</td>
</tr>
</tbody>
</table>
7.1 Introduction to the Performance Management Plan or M&E Plan

Often the PMP or M&E plan (henceforth referred to as the PMP) is limited to a set of indicators table. This is not the full PMP but a piece of it. The PMP as a strategy document which details the “why” and the “how” and the “by whom” of M&E. As such, the document should be a more comprehensive document containing a narrative in addition to tables and graphs. It should tell the whole story. Although the M&E specialist is likely to be the lead on putting the PMP together, the close collaboration of the other members of the project team is essential.

A full PMP is made up of:
- Project Logical Framework
- Identification of critical assumption
- Monitoring strategy
- Evaluation strategy
- Special studies
- Baseline data and targets
- Performance management task schedule
- Indicator Performance Tracking Table (IPTT)
- Performance Indicator Reference Sheets (PIRS)
- Budget

7.2 Components of a PMP or M&E Plan

This is a comprehensive document that should cover all aspects of the data system in such a way as to allow anyone to pick it up and know exactly what needs to happen when. The main sections of the PMP are:

- **Project Logical Framework Analysis.** Include the diagrammatic representation of the Log Frame as well as a narrative explaining the causes and effects relationships that are depicted by the Log Frame, noting appropriate references from past projects and applied research that can justify the causality underlying the log frame (see section 6.2).

- **Identify Critical Assumption.** Critical assumptions are a general condition under which the development hypothesis or strategy for achieving the objective will hold true. Critical assumptions are outside the control or influence of the organization and its partners (i.e., they are not results), but they reflect conditions likely to affect the achievement of results in the RF.

- **Monitoring Strategy.** How will performance be monitored? What activities are proposed to complement indicators tracking? Who will be responsible for data collection, data summary if multiple sites, data entry, data retrieval, reviewing data quality, reporting and disseminating findings? Describe the time line associated with the data flow from frontline collector to program users. *(The data collection/reporting tools can be added to the PMP as an appendix or in a standalone document. But they should be made available.)*
- **Evaluation Strategy.** What are the main evaluation questions that could be derived from the program logic and proposed strategy? What assumptions could be tested and add to the body of knowledge of what works and what does not work? Even when Africare is not responsible for the evaluation, as implementer, it is well positioned to guide and influence the evaluation process and procedures. Provide a schedule for planned evaluations, assessments and special studies during the course of the project cycle. Keep in mind the issues brought up in Module 3 and pay particular attention to the size of the budget and the scope of any evaluations that are planned; they must be in proper proportion to one another.

- **Special studies.** As appropriate, describe any other studies, such as a needs assessment, results analysis and cost-benefit or a beneficiary analysis that was done during the pre-award phase to justify the proposal (see section Error! Reference source not found.), and that are proposed to be carried out during the life of the project.

- **Baseline data and targets.** Discuss plan for collecting baseline data or obtaining it from secondary sources. Provide a timeline and discuss any data issues. Describe how target values have been established and include detailed calculations as necessary in an annex.

- **Performance Management Task Schedule.** Outline the schedule for data collection, review of the data, and assessment of the quality of the data, as well as the office, team, or individual, responsible for ensuring data are available and on schedule. This is the place to develop the data flow chart and diagram. Also, be sure and include in this section a description of the planned Data Quality Assessments.

- **Indicator Performance Tracking Tables.** Include all key information on the indicators that have been identified to monitor progress and performance of the project. Key information includes: Baseline values, Targets, Sources of the data, Periodicity of collection, and who is responsible for the variable (sometimes it is a partner and not Africare that is responsible for one portion of indicators or another.)

- **Performance Indicator Reference Sheets (PIRS).** Contains detailed definition of each performance indicator including: source, method, frequency, how the performance data will be analyzed, and how it will be reported, reviewed, and used to inform decisions. (See Section 6.3 below).

- **Budget.** M&E activities come with a cost. Include a detailed budget by line item, with unit cost where relevant as well as total cost, so that updating can be done easily.

### 7.3 Performance Indicator Reference Sheet

In addition to the standard indicator tables, it is important to develop the “indicator reference sheet”. An indicator reference sheet is a single indicator record form and there should be one for every indicator. It should include all the relevant information about an indicator, past, present and planned. It should provide a detailed definition of the indicator, how it is measured, by whom and where, at what frequency is the data to be collected and how is it to be reported, since there is typically a difference in time and manner between the collection of the raw data and calculation and reporting of the
indicator. The reference sheet should also indicate the baseline and targets for the indicator and how they were calculated.

It may be necessary to refer to another document, in which case, the full reference should be included along with an indication of where it is stored physically or digitally. These reference sheets should be named and filed in a way easily accessible for all “users” of those indicators. Remember that projects are implemented over a period of time long enough for memory to fade and people to come and go.

There are different templates for the indicator reference sheet, included in Annex 9.5 of this document is a template that includes the component pieces that should be included in the Indicator Reference Sheets. If a particular piece of information is not relevant for an indicator or is not yet available for any reason (the baseline survey is not yet completed or we are still waiting for the Ministry of Agriculture to provide details on how yields are calculated), it should be noted on the sheet. These are internal operational tools and need to be useful for people who are collecting the information, as well as those who are using the information.

**REFERENCES AND FURTHER READINGS**


*Note: We have found that the direct link to the USAID TIPS documents did not always work. If you encounter a problem, the following website has links to all of them:* [http://www.innonet.org/resources/node/636](http://www.innonet.org/resources/node/636)
MODULE 8 PUTTING THE PMP TO WORK

PURPOSE OF THE MODULE

Once the M&E plan is in place, it is important to carefully plan its implementation in order to maximize its usefulness for management and decision-making. Among constraints to an effective M&E system are (1) the quality of data, and (2) the delay between the time the data is collected and when it can be used. This delay often makes the retrieved information obsolete or overtaken by events. There are no magic bullets to resolve these problems, but there are tools that can help in reducing the time to process and make available the data and improve quality of the data used for monitoring purposes.

This module will provide as many tools as possible to facilitate the work of managing the M&E system during implementation and improve the quality of the products. Developing project specific tools takes time and diligence and usually the collaboration of other team members. However they are essential for effective monitoring, with the focus set on analyzing and using the information. The objective of setting up these tools should be to ensure that the reporting function of M&E does not become an end in itself, but a means to a functioning evidence-based project management system.

CONTENT OF THE MODULE

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 THE INDICATOR PERFORMANCE TRACKING TABLE</td>
<td>47</td>
</tr>
<tr>
<td>8.2 DEVELOPING OTHER STANDARD FORMS</td>
<td>47</td>
</tr>
<tr>
<td>8.2.1 Data Flow Diagram</td>
<td>47</td>
</tr>
<tr>
<td>8.2.2 Data collection forms</td>
<td>49</td>
</tr>
<tr>
<td>8.2.3 Reporting forms</td>
<td>49</td>
</tr>
<tr>
<td>8.3 DATA MANAGEMENT SYSTEMS</td>
<td>49</td>
</tr>
<tr>
<td>8.4 DATA QUALITY MANAGEMENT PROTOCOLS</td>
<td>50</td>
</tr>
<tr>
<td>8.4.1 Data quality insurance plan</td>
<td>54</td>
</tr>
<tr>
<td>8.4.2 Internal data quality assessments</td>
<td>54</td>
</tr>
<tr>
<td>8.4.3 External data quality assessments</td>
<td>55</td>
</tr>
<tr>
<td>8.4.4 The feedback loop</td>
<td>55</td>
</tr>
<tr>
<td>8.5 CALENDAR FOR REVIEW OF THE IPTT AND FEEDBACK LOOP</td>
<td>55</td>
</tr>
<tr>
<td>8.6 PERIODIC REVIEWS OF THE M&amp;E PLAN</td>
<td>56</td>
</tr>
</tbody>
</table>
8.1 The Indicator Performance Tracking Table

The Indicator Performance Tracking Table, or IPTT, is the main tool for management at Africare to track progress and performance across its portfolio and every project is required to develop an IPTT using a standard format. The IPTT is an Excel spreadsheet with one section containing a fixed set of information – the list of indicators being tracked and their intrinsic characteristics – and another part, with information that is updated every quarter. Step by step guidelines are available on how to set up the table at the onset, with the project-specific indicator information, and how to fill in the table with the quarterly information. Developing a PMP and indicator reference sheets should be done prior to putting together the IPTT table, which would then be a straightforward exercise.

To the extent possible, format used for reporting to donors or for communication purposes should be adapted from the IPTT. This helps to avoid duplication of efforts, reduce risks of confusion and contradiction, and in general standardize the message and format used to convey what has been achieved. A standard for reporting documentation will create smoother data coherence between donor organizations and Africare, allowing information to be more effectively communicated.

8.2 Developing Other Standard Forms

A large component of the M&E specialist’s work during implementation is collecting, verifying and reporting data. Developing standard forms and templates for regular tasks that are incumbent to the M&E team may take some effort up front, but saves time later on, in addition to reducing chances of error and confusion.

8.2.1 Data Flow Diagram

A data flow diagram is a map of the entire M&E reporting system, from the points of data collection to the various reports. It depicts graphically where the data comes from, who is responsible for the different actions, and what data feed into different reports. This kind of diagram is particularly useful when different entities are involved in collecting, processing and analyzing the data, as well as for developing a calendar for data collection and transmission to allow for submitting reports on time.

EXAMPLE

Below is an example of a data flow diagram for a CRS project in the Democratic Republic of Congo (DRC) where targeted farmers are trained to use no till agriculture (NTA).
This diagram shows where the data is first collected and how it flows to the Quarterly Report and Six-month report for the foundation. Each box in the diagram corresponds to either a data gathering activity (with the name of the form on top of the box and the responsible entity at the bottom, in italics) or a report format. The heavy borders contain related set of activities. For instance the larger box on top includes the forms used by extension agents and farmer groups at the demonstration sites, while the larger box on the left contains the reports completed by the implementing staff (CRS and consultants).

Data flow diagrams can also be developed for one or a few related indicators. In this case, the map would trace the various steps data goes through from its point of collection to a final reporting table and is useful for indicators that follow complex trails, such as when the data is transferred several times, different points of collection need to be synchronized, or the same indicator is used in different reports with different time and format requirements. A data flow diagram will also help in identifying data quality vulnerabilities and putting together a calendar to ensure reports are submitted on time (see section 8.4).

Note that a description of the data flow is part of the “Performance Management Task Schedule” in the PMP (see section 7.3). If the flow is relatively straightforward, it can be incorporated into one diagram. For more complex trails, it is advised to develop a separate diagram and attach for each separate trial of data.
8.2.2 Data collection forms

A large amount of monitoring data is collected directly on implementation sites. Whether the information is collected using pen and paper or can be directly entered into a computer, developing data collection templates helps in standardizing data entry and reporting format and is an important step in controlling for the quality of the data. The form should state clearly on top the name and definition of the indicator and other characteristics, such as frequency or disaggregation, that are relevant to the data collection process. The form should identify who is collecting the information and where and when the operation took place. The form should be simple and straightforward to fill and should request only the information directly relevant to the indicator. There should be room on the form for the data collector to add his/her comments, which should be encouraged. It is easier to ignore non-informative comments than to try to retrieve it later on. (See section 2.4 for more discussion on data collection forms and tools.)

8.2.3 Reporting forms

One of the main responsibilities of the M&E specialist is reporting, which may take different forms depending on the audience and purpose. Developing templates for the various reports is recommended since we want to minimize the delay between the time when the data is collected and when it is reported. Most donors will have reporting requirements, which should be used to develop the templates for the M&E data. Usually reports require other information in addition to M&E data, in which case, the whole team should work together to develop the reporting templates.

The main M&E reporting tool is the indicator performance tracking table, or IPTT, that all country offices must fill on a quarterly basis for reporting to headquarters. There is detailed guidance on how to format the IPTT and how to fill it (see section 7.2 and example in Appendix 10.6). To the extent possible, other report templates should mirror the format of the IPTT to facilitate and expedite the reporting process.

8.3 Data Management Systems

M&E data should be handled and stored electronically through a data management program. There are different programs that are available and the choice should be based on the volume of data to be collected and complexity of treatment. Data can be stored in Excel spreadsheets, which allow for simple manipulations and graphical representation. More sophisticated programs offer more functions and can handle larger volume of data, but are usually more complicated to set up and will cost more in time and material.

It is important to plan how particular data will be entered, stored, and analyzed before it is collected. The most appropriate platform depends on the particular data and its use. We present a checklist of some of the characteristics of the data that need to be taken into account when making these decisions about entry and storage. It is usually the case that more than one platform will be required for M&E data. (See section 2.5 for further discussion of electronic management and storage of data.)
CHECKLIST

- How many “variables” and how many “observations” per variable will we have? Most quantitative surveys are easily categorized as large datasets, but in some cases, apparently simple indicators can end up generating large amounts of observations, especially if the project intervenes in multiple sites.

- Are both numerator and denominator data being captured? Sometimes the indicator to be reported is a percentage (e.g., % of adult women engaged in IGA, Numerator is the number engaged and the denominator is the total number of women eligible to be engaged in IGA). In order to report the percent you need both numbers.

- How much analysis and manipulation will be required once the data is entered? Is the data generated quantitative or qualitative in nature? If the data is qualitative, how will it be coded and sorted?

- Does the data contain personal information and confidentiality needs to be protected?

- How many people need to have access to the data and for what purpose (reading only or analysis as well)?

- How long does the data need to be kept?

8.4 Data quality management protocols

We collect regular data as part of the monitoring process so that it can be used for informed, evidence-based decision making. Not all M&E data will be of the same quality and the quality of the same data can vary over time. It may not be a problem if the issues with the data are known and its quality still deemed sufficient for its purpose. The quality of data is determined based on five criteria: validity, reliability, precision, integrity, and timeliness. These five criteria are described below with examples adapted from those presented at the Africare M&E Workshop that was held in Maputo in August 2012.

Validity

Validity refers to the extent to which the data actually measures what it is intended to measure. Validity is linked with the issue of bias in the data and measurement error (representativeness).

EXAMPLES

Result: Poverty of vulnerable communities in conflict region reduced
Indicator: Number of people living in poverty
Source: Government statistics office
There is a problem of validity here because although the government poverty statistics were disaggregated enough to be representative at the level of the communities targeted by the project, it was found that internally displaced people (IDPs) were not included in the sample. The indicator as measured is therefore biased and not a valid measure of the result.

| Result: Modern sanitation practices improved |
| Indicator: Number of residents in targeted villages who report using “clean household” practices |
| Source: Door-to-door survey conducted three times a year |

In this case, the problem found was that many people in the area worked long hours in the fields during the harvest season and were not included as respondents in the surveys. This is an issue of representativeness.

**Additional examples of challenges to validity**

- An evaluation designed to assess the impact of nutrition education on weaning practices is valid if actual weaning practices are observed. An evaluation that relies only on mothers’ reports may find that it is measuring what mothers know rather than what they do.
- An instrument to measure self-esteem in one country may not be valid in another culture.
- Our understandings of validity change over time. It is still debated whether I.Q. tests are a valid measure of intelligence. Long ago, measurements of the skull were thought to be a valid measure of intelligence.

**Reliability**

Reliability refers to the consistency and objectivity of the data collection and analysis methods over time. In other words, does the same procedure of data collection and analysis yield the same results each time it is performed?

**EXAMPLE**

| Result: Employment opportunities for targeted sectors expanded |
| Indicator: Number of people employed by enterprises targeted by the project |
| Source: Structured interviews with targeted enterprises, as reported by three implementing partners. |

This indicator was found not to be reliable because the different implementing partners were using different definitions for “employee”:
• Partner 1 defined an employee as any person receiving wages from the enterprise
• Partner 2 defined an employee as someone working full time for the enterprise
• Partner 3 defined an employee as someone who worked at least 25 hours a week every week

How to improve reliability? When constructing reliable data collection instruments:
• Ensure that questions and the methodology are clear
• Use explicit definitions of terms
• Use already tested and proven questioning methods.

Additional examples of reliable measures
• A reliable reading comprehension test to measure children’s level of competency in English would be one that has the same results from one week to the next, so long as there has been no instruction in the intervening period.
• A ruler is a reliable measure of length.

Examples of measures that may be unreliable
• A questionnaire to measure self-esteem may not be reliable if it is administered to people who have just experienced either success or failure.
• Asking the question “Have you been tested for HIV?” may not yield reliable data because some people may answer truthfully on this sensitive topic and some may not.
• The method of dietary recall to measure food consumption is only as reliable as each respondent’s memory.

Precision

Precision refers to the level of detail or exactness of the data such that it provides sufficient information for its purpose (assessment of performance and make management decision).

**EXAMPLE**

<table>
<thead>
<tr>
<th>Result: Civil Society Organization (CSO) representation of citizen interests at national level increased</th>
<th>Indicator: Average score of USAID-assisted CSOs on the CSO Advocacy Index</th>
<th>Source: Ratings made by the implementer after interviews with each CSO</th>
</tr>
</thead>
</table>

The average scores were reported with different levels of precision from one year to the next:
• 2.55 in Y1
• 3 in Y2
• 3.000 in Y3

The magnitude of the changes between years could not be ascertained because the level of precision in the reporting of the score, through the number of decimal points, varied from year to year. Note that the precision depends on the scale and the size of the variations.
between observations. With more precision, the magnitude of the changes between each year can be more easily captured. Here is the same data, presented with more precision, allowing more insight into the impact of the project:

- 2.554 in Y1
- 2.743 in Y2
- 3.000 in Y3

**Integrity**

Integrity refers to the possibility that the data can be manipulated improperly. This can happen because of faulty data entry protocols, making them vulnerable to transcription errors, or because there are incentives for people collecting or transmitting the data to manipulate the results for political or personal gain.

**EXAMPLE**

<table>
<thead>
<tr>
<th>Result:</th>
<th>Increased agricultural productivity in irrigated perimeters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator:</td>
<td>Average rice yields in irrigated perimeters</td>
</tr>
<tr>
<td>Source:</td>
<td>Yields reported by the Ministry of Agriculture</td>
</tr>
</tbody>
</table>

The problem found here was that the yields reported by the Ministry of Agriculture were systematically higher than available third party measurements. Extension agents and the ministry of agriculture staff had an incentive to inflate the data to demonstrate results.

**Timeliness**

Timeliness refers to whether the data is sufficiently current and up to date to be useful for management decisions.

**EXAMPLES**

<table>
<thead>
<tr>
<th>Result:</th>
<th>Use of modern contraceptives in targeted population increased</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator:</td>
<td>Number of married women of reproductive age reporting using modern contraceptives (CPR)</td>
</tr>
<tr>
<td>Source:</td>
<td>DHS Survey</td>
</tr>
</tbody>
</table>

The problem found here was that of frequency. Because the DHS survey is conducted only every 5 years, results were not obtained on time to inform management on the impact of the interventions.

<table>
<thead>
<tr>
<th>Result:</th>
<th>Primary school attrition in targeted region reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator:</td>
<td>Rate of student attrition for years 1 and 2 at targeted schools</td>
</tr>
<tr>
<td>Source:</td>
<td>Enrollment analysis report from the Ministry of Education</td>
</tr>
</tbody>
</table>
The problem faced here was that the Ministry published in July 2002 enrollment figures for the school year 2000 – 2001 and thus the data was not current and could not be used to assess the impact of the interventions.

8.4.1 Data quality assurance plan

The quality of the data generated by a program and used for management decision is paramount for everyone and should be everyone’s responsibility. In fact, when project team members look only to the M&E specialist to ascertain the quality of the data they miss out on input from project staff who know the project well and are often best positioned to detect early problems with the data. We recommend developing a “Data Quality Assurance Plan” early in the implementation of a project for continuously assessing and documenting the quality of the data that will be generated. (See section 2.6 for further discussion of DQA.)

CHECKLIST of what a Data Quality Assurance plan should include

- An analysis, indicator by indicator, of known or potential data quality issues
- Proposed approaches to assess and monitor quality, including external and internal methods
- A calendar and timeline for project implementation, benchmarks, and assessments
- A system for easy documentation of findings, actions taken, and follow-up reviews that everyone can access

The analysis of data quality issues will necessarily need to be revised as more specific information is obtained on the data and collection procedures. The proposed approaches to assess and monitor quality and the calendar should also be updated as appropriate.

8.4.2 Internal data quality assessments

Internal tools of data quality assessments are measures that the project team can implement to periodically review the data for its quality and usefulness. These include:

- Quarterly review of the indicator data by the project team to report anomalies, such as breaks in trends or suspicious large changes
- Sampling of raw data to independently generate the indicator
- Follow-up training sessions on data collection protocols with cross-examination of how procedures are effectively done
- Visit of project sites to physically verify that what is reported has occurred. When a project site is inaccessible obtain photographs with date/time stamps.
- Comparison of central office records with field office records
- Audits of financial information
- Feedback from users of the data
8.4.3 External data quality assessments

External methods are formal, comprehensive reviews of data quality that include most or all indicators for a project. Formal Data Quality Assessments (DQAs) are technically more rigorous and more costly. The scope of an external DQA should be to assess the data generated against all five standards and also examine the M&E system as a whole to see how it could be made more effective or efficient. DQAs are important for any project to undertake at some point because they are more objective and usually generate a set of specific recommendations that are actionable. How often should formal data quality assessments (DQAs) be conducted depends on a number of things, including the amount of data generated by a project and the complexity of the collection procedures. USAID requires missions to conduct a DQA every 3 years for the data reported to Washington.

8.4.4 The feedback loop

Often data quality issues arise because of mistakes staff are making without realizing it. Maintaining a good feedback loop will help to identify the sources of quality issues early, thus addressing any issues efficiently.

**EXAMPLES of feedback loop actions**

- Bringing the project team together to review the quarterly IPTT before it is finalized so that issues with inconsistent results can be identified.
- Organizing follow-up training sessions bringing people responsible for collecting, storing, or transmitting the data can be helpful to make sure procedures are fully and consistently understood.
- Keeping field agents responsible for implementing the project informed of what results indicators are revealing so as to make them aware of the importance of the data they are generating.
- Comparing results from similar projects implemented in different countries can show unsuspected differences in procedures for collecting data or reporting on an indicator.

8.5 Calendar for Review of the IPTT and Feedback Loop

It is useful to establish a calendar for the country team to review and comment on the IPTT before it is finalized. This review should take place ahead of the PPR, so that any questions raised about the data, can be addressed and the IPTT corrected if necessary. Project team members who know the details of how and when activities are implemented are well positioned to detect problems with the indicators. Rarely do M&E specialists have that same level of understanding about the implementation, especially across the portfolio.
8.6 Periodic reviews of the M&E plan

As the implementation of a project progresses, changes to the indicator characteristics, data collection methods, and calendar will necessarily occur. New information is acquired and the test of implementation requires constant adjustments. It is important to document any change that seems necessary, whether it happens immediately or not. Again, memory fades and people come and go. A template can be designed to record proposed changes and a folder created to save any relevant documents, email exchanges, memos, etc. Some changes may be relatively straightforward, such as clarifying the definition of an indicator, but others, such as dropping or adding an indicator or new evidence suggesting that the program Logical Framework should be revisited, require more elaborate justification and documentation. It is good practice to include periodic updates of the PMP in the work plan, where these changes and updates can be formalized and fully documented. It does however take time and effort to revise the PMP, so it is better to do it at certain intervals when there is time to devote to the task. Revisions to the PMP typically require consultation and validation by project team members, implementing partners and other stakeholders, which should be taken into account in the planning.

REFERENCES AND RECOMMENDED READINGS


*Note: We have found that the direct link to the USAID TIPS documents did not always work. If you encounter a problem, the following website has links to all of them: http://www.innonet.org/resources/node/636*
MODULE 9  METHODS AND TOOLS FOR DATA COLLECTION

PURPOSE OF THE MODULE

This module focuses on data: what is data and what are the different techniques available to M&E for collecting it. Some data collection can be set up as part of the project implementation mechanism, while others require specialized tools and most likely hiring a firm or consultant to do the work. The module will describe tools to assess what kind of data is best suited to different situations and how to obtain it.

CONTENT OF THE MODULE

| 9.1 | TYPES OF DATA | 58 |
| 9.1.1 | Primary versus secondary data | 58 |
| 9.1.2 | Quantitative versus qualitative data | 58 |
| 9.2 | METHODS OF DATA COLLECTION | 59 |
| 9.3 | QUALITATIVE METHODS | 59 |
| 9.3.1 | Interviews | 60 |
| 9.3.2 | Focus Groups | 60 |
| 9.3.3 | Direct observation | 61 |
| 9.3.4 | Mystery Clients | 61 |
| 9.3.5 | Hearsay Ethnography | 62 |
| 9.3.6 | What to include in a qualitative Scope of Work | 62 |
| 9.4 | QUANTITATIVE DATA COLLECTION TOOLS | 62 |
| 9.5 | SAMPLING | 63 |
| 9.5.1 | Non-probability sampling | 64 |
| 9.5.2 | Probability sampling | 64 |
| 9.5.3 | Cluster sampling strategy | 65 |
| 9.5.4 | Sample size | 65 |
| 9.5.5 | Stratification | 66 |
| 9.6 | DATA QUALITY | 66 |
| 9.7 | DATA ANALYSIS | 67 |
| 9.8 | WHAT TO INCLUDE IN A SURVEY Scope of Work (SOW) | 68 |
9.1 Types of data

Data can be categorized between primary and secondary and between qualitative and quantitative. It is useful to understand the distinctions because these categories of data differ as to how they are collected, how they can be analyzed, and how they should be used and validated.

9.1.1 Primary versus secondary data

Primary data is data that is collected directly by the user for a specific purpose. It is original data. Primary data can be collected through a structured survey, direct observations, focus group discussions, interviews or other direct methods. Baseline data that country offices collect at the start of implementation is an example of primary data, and so is project-level monitoring data, such as “number of farmers trained by the project” or “percent of women who say they have attended a follow-up information meeting”.

Secondary data is data that is used subsequently by someone who did not collect it. The data was collected primarily for another purpose. Sources of secondary data include institutional records, macroeconomic statistics, and literature reviews. Data collected by health clinics for their management purposes and used by an Africare health project for monitoring purposes are considered secondary data.

9.1.2 Quantitative versus qualitative data

Quantitative data is a numerical measurement (can be meaningfully expressed as a number), associated with a scale measure (an ordinal, interval or ratio scale). Quantitative data can be analyzed through statistical methods. We distinguish further between nominal and ordinal categories. Nominal categories do not have a natural ordering, such as gender or ethnicity. Ordinal categories have a natural ordering, such when we describe an attitude (strongly agree, agree, disagree, strongly disagree) or satisfaction level (highly satisfied, satisfied, dissatisfied).

Qualitative data methods help build an in-depth picture among a relatively small sample of people on a specific issue. They reveal in more detail how people perceive their own situation and problems, why and what their priorities are. Qualitative methods are flexible. Questions are asked in an open-ended way and the findings are analysed as data is collected. When qualitative methods are used systematically, the findings are as reliable and objective as those produced by quantitative methods. With qualitative data we are trying to uncover why something is happening and we look to find out information on feelings, impressions and opinions that are often difficult to obtain through structured quantifiable questionnaires.

9 Note that not all numbers are quantitative data. Phone numbers for instance are made of numbers but cannot be “scaled”, added or subtracted for instance.
9.2 Methods of data collection

There are different methods to collect primary data for monitoring purposes: quantitative and qualitative methods. Quantitative methods are mainly structured surveys with discrete responses and direct measurements. Qualitative methods rely on open-ended questions and are more free-flowing in approach. They include such methods as focus group discussions, interviews, and direct observation.

The following is a list of things to take into consideration when selecting a methodology:

<table>
<thead>
<tr>
<th>Consideration guiding the choice of method(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose and information required</strong></td>
</tr>
<tr>
<td>▪ The decisions that users will make based on the results of data collected and the corresponding questions guiding data collection e.g. certain questions require qualitative or quantitative data.</td>
</tr>
<tr>
<td><strong>End users’ preferences</strong></td>
</tr>
<tr>
<td>▪ What types of data do different stakeholders and partners find most convincing? Where these differ, whose preferences will be given priority?</td>
</tr>
<tr>
<td><strong>Information available</strong></td>
</tr>
<tr>
<td>▪ How has existing information been gathered? Is there opportunity to use consistent methodologies to build on and compare to existing data?</td>
</tr>
<tr>
<td>▪ Methods used in the past will point to practical limitations and possibilities, e.g. what sampling frames are available; what methods are familiar to local staff.</td>
</tr>
<tr>
<td><strong>Level of precision required</strong></td>
</tr>
<tr>
<td>▪ What level of precision is needed for decisions to be made?</td>
</tr>
<tr>
<td><strong>Coverage/depth required</strong></td>
</tr>
<tr>
<td>▪ What is the geographical coverage required for data collection? National?</td>
</tr>
<tr>
<td>▪ Which groups with different status and power within community (e.g. gender, wealth, age, ethnicity).</td>
</tr>
<tr>
<td><strong>Time frame</strong></td>
</tr>
<tr>
<td>▪ When are results needed by decision-makers? Does this timeframe exclude any methods?</td>
</tr>
<tr>
<td><strong>Human resources</strong></td>
</tr>
<tr>
<td>▪ What skills and expertise are readily available?</td>
</tr>
<tr>
<td>▪ Can additional human resources be recruited, contracted if needed?</td>
</tr>
<tr>
<td>▪ Is additional training will be required?</td>
</tr>
<tr>
<td><strong>Financial resources</strong></td>
</tr>
<tr>
<td>▪ How much will a data collection activity cost?</td>
</tr>
<tr>
<td>▪ Consider also the opportunity-cost, the impact on other activities of resources used for this data collection (e.g. use of vehicles, interpreters).</td>
</tr>
</tbody>
</table>

Adapted from: http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/CHAP_3.HTM and Danida and Roche

9.3 Qualitative methods

Qualitative methods are best used to collect in-depth information from a small number of people. They are helpful in providing context and describing specific circumstances. Their fields of enquiry are:
Beliefs and attitude, social norms and values, and social interactions (such as in terms of gender, ethnicity, income)
• Institutional or social dynamics (such as cooperation and conflict)
• Attitude and practices related to services delivery (such as health, education, food aid)

**CHECKLIST – when to use qualitative methods**

- Quantitative data reveal something that cannot be explained
- You need to verify if findings from an older survey or from secondary data can apply to your context
- You want to know what people think about an issue or a new proposal, or you want to know what their priorities are
- You need to identify or validate topics and questions for a quantitative survey
- You need to understand a certain behavior sanctioned by the community
- You need to quickly obtain an answer at low cost and representativeness is not a concern

The qualitative methods most commonly used for monitoring and evaluation are: interviews, focus group discussions, rapid appraisals, and direct observation.

**9.3.1 Interviews**

There are different kinds of interviews depending on whether the interviewer uses a checklist of topics and structured guide, or if the interview is conducted as an informal conversation. How respondents are selected is important (see section 9.5.1). Individuals can be picked randomly, opportunistically (whoever comes first), or purposively (because they have a particular knowledge or interest in the topic and can serve as key informants). The purpose of individual interviews is to gain insight into the experience of an individual on a particular subject. The individual is encouraged to provide explanation and rationale and expand on related topics. As such, individual interviews can be time consuming and must be conducted by skilled interviewers. Ideally, interviews are recorded (with permission) and the interviewer takes notes, so that the information can later be analyzed for feedback and direction.

**9.3.2 Focus Groups**

Focus group discussions are also conducted around a topic or a small number of related topics and are facilitated through the use of a structured Field Guide containing multiple key open-ended questions. The intention of the guide is to support participants in opening up and sharing as much as possible on the topics of interest. Often they share information that is not on the guide and that is an added bonus. Because the interactions between the participants are part of the information sought, it is important to decide in advance how people will be selected and for the interviewer to remain neutral in the
discussion. In fact, the less the interviewer talks and the more the participants talk the better the outcome.

These group discussions are designed to reveal the experiences and perceptions of individuals in a social context, where interpersonal dynamics are at play. Focus group discussions should be carefully planned and bring together a small number of pre-selected people (generally 6-10 people). Participants of focus groups can also be selected randomly, purposively, or opportunistically (see section 9.5.1). It is recommended to select participants in such a way that they will feel at ease with each other and comfortable speaking up. This implies that groups will be homogeneous and be selected based on their social and individual similarities, such as their sex, age, ethnicity, income group characteristics and the like. Each type of homogeneous group should be run at least 2 or 3 times with new participants in order to smooth out the effect of “outlier” groups. Focus group discussions should be recorded to the extent that participants agree and the discussion should be transcribed and then analyzed to inform the assessment.

9.3.3 Direct observation

Direct observation simply means using what you observe to record information. We constantly make observations, but in order to be useful as a data collection tool, observations must be organized and purposeful. In many instances, the data used to inform indicators, at the input or output level in particular, is from direct observation: a training session has been held, community food crop storage facilities are rehabilitated, etc. Direct observation can also be used to gain further insight on a specific topic. For instance, we may be interested in observing how a town market functions and set out to observe: how products are brought to the market and at what time; how do large and small retailers interact; and how flows of merchandises vary during the day. Direct observation as a method of data collection should be planned and organized with templates to record and analyze the information.

9.3.4 Mystery Clients

The ‘Mystery Clients’ concept is an important research technique that serves as a performance improvement tool. It helps to monitor and evaluate service delivery programs, health facilities and providers, through providing information on good practices as well as essential service delivery elements that may be overlooked during routine clinic tasks and activities.

Mystery clients are trained people who visit health facilities in the assumed role of clients, and then report on their experience. Using mystery clients/customers is very useful for collecting information about health facilities and providers, but should be conducted with the voluntary consent of the participating service providers and clinics. When the findings are used to assess the quality of services, this technique serves the interests of both clients and the program by identifying the ways that service provision can be improved to attract more customers.
9.3.5  Hearsay Ethnography

Hearsay ethnography is a method that is used to help get to what your target group is saying when you are not around. This method can help you in finding out how your population is relating to your intervention or how they are learning, not learning, understanding, not understanding or changing or not changing behaviors. This is ethnography so it does not allow you to generalize to populations beyond your own target group and should be used to give you a 'sense' of what is happening in your target population. It can be used to help you to find out whether or not your target group is making the changes in knowledge, attitudes, social norms and behaviors that you are hoping they are changing.

This method requires that you recruit several members of your target population to write down, at the end of each day, what was said in conversations they overheard or participated in as they went about their normal activities. Although originally developed to augment data collected by demographers studying AIDS in sub-Saharan Africa, hearsay ethnography turns out to have broader methodological and theoretical advantages. Unlike surveys, focus groups, or in-depth interviews, hearsay ethnography captures spontaneous conversation, rather than responses to researchers’ questions or probes. Local residents can also gather data from many more locations and social contexts than a traditional ethnographer could do, including gender-segregated settings and those to which an outsider would not have access. More important, hearsay ethnography reveals the varied personal and collective agendas that lead people to confront AIDS, and it captures the ongoing shared enterprise of constituting and revising cultural understandings.

9.3.6  What to include in a qualitative Scope of Work

A qualitative data collection activity should be carefully planned and appropriate tools developed and tested. These should be first developed into a standard protocol (see template in annex) and then can be used to set up a scope of work (SOW), whether the work is to be done in-house or by a consultant. The SOW will include a description of the issues and topics or questions requiring further investigation, how the participants or interviewees will be selected, a work plan (training schedule for interviewers, pilot testing of tools and procedures, location of interviews, debriefing of interviewers), an analysis plan, a description of the tools, and a dissemination and use plan. Specific tools to be developed depend on the methods chosen, but would include interview guides and forms to take notes. Specialized software to code the information should be used if relatively large amounts of data are collected and need to be rigorously analyzed. Key, quality information is too often negated by informal, subjective analysis and anecdotal presentation.

9.4  Quantitative data collection tools

Aside from a direct measurement (counting the number of people attending a training session, recording the number of visits to a health care center), the main tool for primary quantitative data collection available to M&E is the statistical survey. The statistical
A survey is used to collect specific information from a sample of individuals that are representative of a population.\textsuperscript{10} Survey methods are designed to obtain reliable, objective data that can be replicated and generalized. It is these two characteristics, the replicability and the ability to generalize, that make the survey such a powerful tool. These also come with constraints in the rigor of the methods necessary in planning and implementing a survey.

A survey can be described in terms of its population of interest: household, enterprise, or agricultural for instance, or the kind of information it seeks to obtain: “household expenditure surveys” or “nutritional surveys”. Some surveys are known for their specific methodologies that have been applied to many countries and over time: the DHS (Demographic and Health Survey) or the LSMS (Living Standard Measurement Survey).

A survey uses a pre-established structured questionnaire to gather data from a sample of individuals. The questions are not open-ended and seek answers from a predetermined set of options or list. Respondents are not expected to provide an explanation for their answers or their views on the questions. Explanations are obtained from the statistical analysis done on the data, where variables are linked and crossed and relationships (correlations and associations) are estimated. A survey is administered on a sample of the population of interest. The population is usually not the whole population, but should be as specifically defined as appropriate and feasible to meet the needs of the programs (the existence of a “sample frame”, or listing, for the population of interest is often a constraint).

\subsection*{9.5 Sampling}

Sampling is a whole area of expertise and even statisticians refer to sampling experts. It is important when considering data collection methods to understand some basic concepts and in particular to distinguish between probability samples and non-probability samples.

\textsuperscript{10} A census collects specific information and uses statistical methods, but it does so on an entire population rather than from a sample. The main purpose of a census is for a central government to obtain an accurate count and descriptive characteristics of the population (such as the stratification by age, gender, and ethnicity, average household size and the like). Censuses are expensive and thus conducted at large time interval, often 10 years or more, and limited to a few questions.
9.5.1 Non-probability sampling

Non-probability samples are samples of respondents that have been selected in a way that prevents researchers from knowing what the individual respondent's probability of being selected for the research is. In an ideal study, all respondents have an equal and non-zero probability of being selected for the study. Non-probability approaches are primarily used in qualitative research (see section 9.3) to identify respondents or participants. In this case, statistical analysis will not be used and the process for selecting respondents is based on criteria other than statistical representativeness. There are two main types of non-probability samples.

- Purposive sampling: respondents or participants are chosen based on known characteristics that make them more likely to provide the information that is sought on the topic of interest (key informant interviews for example)
- Opportunistic/convenience sampling: respondents or participants are selected based on their presence or availability at the moment of conducting the interviews.

9.5.2 Probability sampling

Probability sampling allows for statistical inferences and is the method used in surveys. Probability sampling means that each unit (e.g. individual, household, farm, enterprise) in the population of interest has an equal and known probability to be included in the sample.

A number of rules apply for a sample to be a probability sample, whichever method is used to draw the sample and whether stratification is used or not. In addition, the techniques to calculate the minimum sample size and weights to apply to sample estimates in order to extrapolate results to the population can be complicated. Finally, appropriate rules should be developed for replacing a unit in a sample that needs to be dropped (the unit cannot be found or refuse to answer for instance). In general, it is recommended to always seek proper expertise to ensure that the sampling is done and interpreted correctly.

In order for the criteria of a probability sample to be met, one needs a listing of all the units in the population (the sample frame). This means you need the listing of all the children in the schools to be sampled, all the households in a village, all patients in a clinic, etc... Population census data are commonly used to constitute a sample frame. However, censuses are not always available or an acceptable source of data, especially in developing countries. Problems that may be encountered include the following.

- A census does not exist or is too old
- A recent event in the area of interest has created a demographic shift that is not captured by the last census (i.e. a major flood that has forced people to relocate or the opening of a mine, which has attracted people to the area)
- The census units are different from the study units. Often censuses will provide population numbers and characteristics by administrative units, which do not necessarily correspond to the geographical areas of study. Also, censuses are often
interested in counting the number of individual people, while surveys are often looking at households as the main unit.

- The population of interest may not be counted within the census data (i.e. Sex Workers, men who have sex with men, mobile populations)

If a sample frame must be constructed, a multi-stage approach, or cluster sampling strategy, can be employed.

### 9.5.3 Cluster sampling strategy

A cluster sampling strategy uses a multi-stage approach. To illustrate the procedure, we use an example of a two-stage process, with villages and households as clusters. In the first stage, a listing of villages is either obtained from a census or other government administrative documents, or, if feasible (i.e. the area covered is not too large), it is developed. A sample of villages is chosen randomly from that list, constituting the first stage sampling. In the next stage, a listing is done of the next level unit (households in this example) present in the villages sampled in the first stage. A random sample is then drawn from the household listing.

The number of households selected at the second stage is determined by the sample size calculation (see next section). The number within this sample size selected within each chosen village should be determined by the proportion of all households in the area of interest that are within each village and should not be fewer than ten (10) per village. In the end the proportion of homes in each village that make up the sample will reflect the proportion of homes in each village that make up the entire population; this is called selection of the population proportionate to size. If the number of households in village X equals 60% of all households in the geographic area of interest then the number of households from village X in the sample will also equal 60%. This helps to ensure that the households in each village have the same non-zero probability of being selected for the study.

### 9.5.4 Sample size

The question of the sample size is in fact what the *minimum* sample size should be. The sample size determines the level of statistical significance that can be applied to the statistical results (the interval of confidence). The larger the sample size, the higher the statistical significance of the results will be. Larger samples, however, increase the cost and time of surveys and there is always a point beyond which a larger sample size does not add value. Yet, it is important to ensure that the sample size is large enough. The benefits depend on what the survey will be used for and, in particular, the size of the changes we wish to estimate (between units or over time). For instance, a sample size that yields estimates that are significant at the 90% level are not sufficient to draw conclusions of variations around key parameters less than the 10% level. Note that there are factors other than the size of expected changes that need to be taken into account when calculating the appropriate sample size and proper expertise should be sought to calculate the size of the sample. Examples of sample size calculations can be found in the annex of this document.
9.5.5 Stratification

Stratification is a process of dividing the population into sub-groups that are known to be relevant to the study and drawing random samples within each stratum. Stratification improves the representativeness of the sample since each stratum will be represented (reduces sampling errors). A sample size must be calculated and applied to each stratum of interest. The conditions for stratification are that each sub-group is mutually exclusive and collectively exhaustive. In other words, an individual, or other survey unit, can be assigned to only one sub-group and every unit will be assigned to one of the group. Common examples of strata that are used in surveys are gender, income level, geographical areas, or housing characteristics.

9.6 Data Quality

The data you collect are meaningful only if they are of the highest possible quality. Data quality must be monitored at every single step of the process and such monitoring should not depend on only one person to ensure data quality. Data quality is important because the quality of the data determines the usefulness of the results. This is as true for data that you are collecting for the program as it is for any secondary data that you may be using to help guide your thinking and make decisions.

When using secondary data it is important to get a thorough understanding of who collected the data, how the data was collected, what instruments were used, how interviewers were trained, what the intent of that data collection activity was, and what are the specific qualities and characteristics of the target population of that data. With this information you can better determine if the data meets your needs.

Ask yourself:

- Does this population match our target population(s)?
- Did they use a rigorous study design?
- Did they have a large enough sample size?
- Did they run enough focus groups (5 – 7 per demographic)?
- How well trained were their interviewers?
- Did they get the approval of an Ethics Review Committee?
- What is the reputation of the institution that carried out the data collection?
- Where is it being reported?

This information will help you determine if you want to use secondary data to guide your program. You may find that it is a well-known source and a data set that is collected every few years and is widely accepted as solid and useful information. But don’t negate a data source just because it is not the DHS. Other smaller institutions and agencies may have exactly what you need. Keep your search broad but maintain an eye on the quality of what is collected.

When conducting your own data collection activities there are many ways to ensure data quality. Most of these measures rely on good planning and supervision. The following are some ways that programs might ensure good data quality:
Develop clear goals, objectives, indicators, and research questions
- Plan for data collection and analysis
- Pre-test methods and tools
- Train staff in monitoring and evaluation, and data collection
- Create ownership and belief in data collection among responsible staff
- Incorporate data quality checks at all stages
- Supervisors need to review all forms to make sure that they are complete
- Supervisors need to check to ensure that all answers are clearly written
- Supervisors need to check to see if answers are consistent
- Supervisors need to check to ensure that all figures are tallied correctly
- Check data quality regularly
- Take steps to address identified errors right away
- Document any changes
- Improving the data collection system as necessary

After information has been collected from the field, it is usually entered into a computer. At this stage, more quality checks are necessary because there are common sources of error that arise during data entry.

The following are some common sources of error:

- **Transposition**—an example is when “39” is entered as “93”. Transposition errors are usually caused by typing mistakes.
- **Copying errors**—one example is when “1” is entered as “7”; another is when the number “0” is entered as the letter “O”.
- **Coding errors**—putting in the wrong code. For example, an interview subject circled “1” = Yes, but the coder copied “2” (which = No) during coding.
- **Consistency errors**—consistency errors occur when two or more responses on the same questionnaire are contradictory. For example, if the birth date and age are inconsistent, or a male has stated he is pregnant.
- **Range errors**—Range errors occur when a number lies outside the range of probable or possible values.

### 9.7 Data Analysis

The planning for analysis actually begins months and months before it is even collected. Program managers and planners need to be thinking about how they will use the data and be prepared to collect and analyze it with these uses in mind. The main purposes of data analysis are to first check and see whether you are achieving program objectives and then to summarize data results so that you can decide if changes need to be made to the program. In addition to the direct programmatic uses, the data should be analyzed with specific populations in mind for dissemination. Who needs to see these results? The short list for dissemination is: Funders, community leaders and representatives, partner agencies, and perhaps the media. Once we have collected and cleaned the data, we are sure it is of the highest quality possible, and we know how we are expecting to use the data we are ready for analysis.
Analysis does not mean using a complicated computer analysis package. It means taking the data that you collect and looking at it in comparison to the questions that you need to answer. For example, if what you need to know is if your program is on track, you would look at your program targets and compare them to the actual program performance. This is analysis (e.g., through analysis you find that your program achieved only 10% of its target; now you have to figure out why).

Below are several examples of questions to be answered based on the analysis technique used and the focus of the analysis.

<table>
<thead>
<tr>
<th>Focus of Analysis</th>
<th>Analysis Technique</th>
<th>Questions to be Answered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of program</td>
<td>• Compare actual performance against targets</td>
<td>• Is the program on track?</td>
</tr>
<tr>
<td>performance</td>
<td>• Compare current performance to prior year</td>
<td>• Did we meet our targets? Why or why not?</td>
</tr>
<tr>
<td></td>
<td>• Analyze trends in performance</td>
<td>• How does this period’s performance compare to last period?</td>
</tr>
<tr>
<td></td>
<td>• (All the above techniques can also be by types of services delivered)</td>
<td>• What happened that we did not expect? Are new targets needed?</td>
</tr>
<tr>
<td>Diversity of target</td>
<td>• Comparison between sites or groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity of program to its</td>
<td>Component analyzed:</td>
<td>• Is the program performing functions as it was expected to or is not performing them as</td>
</tr>
<tr>
<td>design</td>
<td>• Training</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Referral system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Services</td>
<td></td>
</tr>
<tr>
<td>Measurement of changes in key</td>
<td>• Comparison of baseline and follow-up population based surveys</td>
<td>• What is the prevalence of key risk behaviors?</td>
</tr>
<tr>
<td>risk behaviors and antecedents</td>
<td>• Comparison of intervention and comparison or control sites</td>
<td>• Have they changed over time? If so, how have they changed?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• What direction are they changing?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Who in the population is making the change?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Are social norms towards the behavior changes supporting or hindering the change?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• After the program, are they more supportive or less or no change?</td>
</tr>
</tbody>
</table>

9.8 What to include in a survey Scope of Work (SOW)

Careful planning and piloting is required before conducting a survey. A detailed protocol must first be developed following the template provided in the Annex. From this protocol a SOW can be drafted if some or most of the steps are to be done in-house. Especially if a solicitation request is put out for a survey, it is very important to include as many details
as possible and be specific about the steps that need to be completed by the consultant. A proposal should answer to each of these components in a clear and coherent manner since both the quality and the cost of the survey will critically depend on each of them.

**CHECKLIST: components of a statistical survey SOW**

- **Purpose of the survey.** What are the questions or hypotheses that need to be investigated or tested? What do we know of the context, population under study, or evidence that exists from other surveys and studies? Is there a prior survey that this survey should be based on?

- **Sampling methodology.** What kind of sample is required (simple random or multi-stage clustering)? What data exist that could be used to constitute the sample frame? Do we have some prior knowledge on factors that should affect the sample size, such as the level of precision necessary and the size of expected change in key variables? Do we have a need for the sample to be stratified (divided into homogenous segments)?

- **Questionnaire development.** What are the topics that should be included in the questionnaire? Is there some limitation in length for the questionnaire to respect? Should there be a different questionnaire for different respondents? Is there a need to interview men and women separately?

- **Pre-test or pilot.** A questionnaire, even if it has been used in the past, should be tested or piloted before it is finalized. The purpose of piloting the questionnaire is to check whether it can be administered as intended, whether the questions are well formulated, how much time it will take to go through all the questions, etc. It is usually recommended to test the questionnaire outside of the sample to avoid any contamination. A revised questionnaire should be submitted following the pre-test.

- **Survey manuals.** Guidelines on how to administer a questionnaire and enter the data should be developed. Should specific guidelines be developed for the supervisors? Guidelines should be specific on what to do in cases where a respondent cannot be found or refuses to answer.

- **Staffing plan.** How and how many field staff will be hired? How many supervisors will be necessary? What experience will be required and what working conditions will be provided (transport, lodging, etc). Do we have specific gender requirements (i.e. should women be interviewed by female enumerators and should men be interviewed by male enumerators)? How are these justified in terms of the proposed implementation plan (calendar and physical deployment)?

- ** Enumerator training.** How extensive should the enumerator training be? Should the enumerators and supervisors be evaluated after the training before the hiring is completed? Can the trainees be used to test the questionnaire?
Data entry. Will the questionnaire be administered with pen and paper or with PDAs? When will the coding templates be developed and who will be responsible for data entry? Will double entry system be used if paper and pen are used in the field? What program will be used to produce the raw dataset? How will the confidentiality of the respondents be protected?

Implementation plan or work plan. A detailed calendar and deployment plan for the implementation of the survey itself should be developed. Sufficient time should be included for data verification, supervisory visits, troubleshooting, etc.

Budget. A detailed budget should be developed, consistent with what is proposed for each step.

REFERENCES AND RECOMMENDED READING

- DHS (Demographic and Health Survey): http://www.measuredhs.com/

Note: We have found that the direct link to the USAID TIPS documents did not always work. If you encounter a problem, the following website has links to all of them: http://www.innonet.org/resources/node/636

---

\[\text{Digital technology such as PDA (Personal Digital Assistant) is increasingly used for administering questionnaires even in remote areas in Africa and elsewhere with great success. PDAs save enormously in terms of time and quality by reducing errors while entering or transferring the data.}\]
# Module 10: Annex

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.1</td>
<td>Glossary</td>
<td>72</td>
</tr>
<tr>
<td>10.2</td>
<td>Sample Size Calculations</td>
<td>75</td>
</tr>
<tr>
<td>10.3</td>
<td>Survey Respondent Consent Forms</td>
<td>76</td>
</tr>
<tr>
<td>10.4</td>
<td>Qualitative Methods Guidance</td>
<td>81</td>
</tr>
<tr>
<td>10.5</td>
<td>Indicator Reference Sheet for the PMP</td>
<td>86</td>
</tr>
<tr>
<td>10.6</td>
<td>Illustrative Indicator Table for the PMP</td>
<td>88</td>
</tr>
<tr>
<td>10.7</td>
<td>Advantages and Disadvantages of Various Data Collection Methods</td>
<td>89</td>
</tr>
<tr>
<td>10.8</td>
<td>Evaluation Research Protocol Template</td>
<td>92</td>
</tr>
</tbody>
</table>
10.1 Glossary

Activity – An activity is a component of a project that contributes to a project purpose. It refers to an award (such as a contract, grant or cooperative agreement), or a component of a project such as training or technical assistance.

Assessment - The examination of the situation prior to an intervention. Involves not only determining what the situation is, but also why it is that way and what opportunities and risks might affect efforts to change the situation. In evaluation literature, assessment is considered a type of evaluation, called an *ex ante evaluation*.

Effectiveness - Effectiveness indicates the extent to which a project or program has achieved its objectives.

Efficiency - Efficiency compares the use of resources (personnel, time and financial resources) with the achieved results (“Input – Output”).

Evaluation - A process that tries to determine as systematically and objectively as possible the worth or significance of an intervention or policy. This judgement is based on common “evaluation criteria” such as relevance, efficiency, effectiveness, impact and sustainability. (See “Linking evaluation criteria and evaluation questions”.) An evaluation should provide information that is credible and useful so lessons learned are taken into account in the decision-making process of program partners and donors.

Note, the word "objectively" is used to indicate the need to achieve a balanced analysis, one that recognises bias and balances the perspectives of different stakeholders (all those interested in and affected by programmes, including beneficiaries as primary stakeholders) by using different sources and methods. Evaluation asks the questions: Are we doing the right thing? And are we doing things right?

Evaluation team - A group of persons responsible for conducting an evaluation. The group composition must be balanced and contain independent experts, representatives of the partners and/or the beneficiaries, and external and local experts. Care must be taken to ensure that the group possesses social as well as technical competences. Special attention is given to ensuring that the members of the evaluation team are able to handle priority cross-sectoral issues (e.g. poverty, gender, and environment).

External evaluation - An evaluation conducted by one or more persons outside the project/program structure.

Focus groups - a means of gathering information through open-ended questioning of a group of people that are members of the affected population or stakeholders in the issue. The focus group method often provides in-depth information but does not produce quantitative data such as the number of families affected or a quantitative assessment of their most critical needs. The focus group method is best used in situations where the service organization has little idea of the problems facing their target population. Information from focus groups can help guide the development of a survey or other quantitative study.

Impact - The long-term, cumulative effect of programs/interventions over time on what they ultimately aim to change, such as a change in HIV infection, AIDS-related morbidity and mortality. Note: Impacts at a population-level are rarely attributable to a single program/intervention, but a specific program/intervention may, together with other programs/interventions, contribute to impacts on a population.
Indicator - An indicator is a variable which helps to identify indirect differences in quality or quantity within a defined period of time. To this end, complex problems are simplified and reduced to an observable dimension.

Input - The resources (know-how, financial, personnel, etc.) required for the planned project output.

Key informant interviews - A key informant is a person (or group of persons) with unique skills or professional background related to the issue/intervention being evaluated, who is knowledgeable about the project participants, or has access to other information of interest.

Logical framework - The process which leads to a planning platform. The various actors within a program set down their visions, jointly analyze the problems to be solved and the available resources and potentials. Based on this they determine the overall aim of a program/project, from which they derive the specific goals, and harmonize their activities and the results to be achieved (output and impacts).

Monitoring - A process of tracking or measuring what is happening. Monitoring begins at the planning stage and is an essential steering element. It is performed in relation to outputs, outcomes, processes, impacts, and the project setting. This includes two kinds of monitoring: Performance monitoring entails measuring progress in relation to an implementation plan for an intervention (i.e. a programmes / activities, strategies, policies) and specific objectives; Situation monitoring entails measuring change in a condition or a set of conditions or lack thereof (e.g., changes in the situation of children and women or changes in the broader country context).

Outcome: Short-term or medium-term effect of an intervention’s outputs, such as a change in knowledge, attitudes, beliefs, behaviors.

Outputs: The results of program/intervention activities; the direct products or deliverables of program/intervention activities, such as the number of HIV counseling sessions completed, the number of people served, the number of condoms distributed.

Overall aim - The overall aim to which a project/program contributes with its results (outcomes), but whose achievement is outside the project’s/program’s sphere of responsibility and is still impacted by other factors.

Participatory research - research that is conducted as an equal partnership between traditionally trained "experts" and members of a community.

Program - A program comprises several projects which pursue a common or similar overall aim. However, a program is more than the sum of its parts.

Project - A project is a planned package of measures to achieve a defined objective within a limited time and geographical area, and within defined financial parameters. The work plan, the mandate, and responsibility for project execution are clearly defined. A project calls for multidisciplinary cooperation between persons and institutions.

Quantitative - With quantitative methods, things are either measured or counted, or questions are asked according to a defined questionnaire so that the answers can be coded and analysed numerically. Quantitative methods:
- Provide a broad view of a whole population based on data from large numbers of people or cases about a limited set of questions

- Can identify major differences in the characteristics of or conditions affecting a population

- Determine whether there is a statistical relationship between a problem and an apparent cause.

**Qualitative** - Qualitative methods help build an in-depth picture among a relatively small sample of people on a specific issue. They reveal in more detail how people perceive their own situation and problems, why and what their priorities are. Qualitative methods are flexible. Questions are asked in an open-ended way and the findings are analysed as data is collected. When qualitative methods are used systematically, the findings are as reliable and objective as those produced by quantitative methods.

**Relevance** - The question of relevance refers to the correlation between a project’s/program’s development objectives and prevailing conditions, opportunities and risks in a particular country. Questions are asked about the extent to which a project helps to meet the existing needs and priorities of the target groups/beneficiaries (are we doing the right thing?).

**Reliability** - A data collection method or instrument is considered reliable if the same result is obtained from using the method on repeated occasions.”

**Research:** A systematic, intensive study intended to increase knowledge or understanding of the subject studied, a systematic study specifically directed toward applying new knowledge to meet a recognized need, or a systematic application of knowledge to the production of useful materials, devices, and systems or methods, including design, development, and improvement of prototypes and new processes to meet specific requirements.

**Results** - Results contain as well “outputs” (direct products and achievements of a program/project) and “outcomes” (their direct and indirect effects).

**Triangulation** - Wherever possible, use multiple sources of data. The different types of triangulation: Data triangulation uses a variety of data sources in a study. For example, interviewing people in different occupations, status positions or belonging to different political parties on one or a series of issues. Investigator triangulation uses several different evaluators or social scientists. Method triangulation uses multiple methods to study a single problem or program, such as interviews, observations, questionnaires or written secondary sources. Theory triangulation uses multiple perspectives to interpret a single set of data.

**Validity** - Validity is the best available approximation to the truth of a given proposition, inference of conclusion. A measurement method or instrument is considered valid if it measures what it intends to measure.
10.2 Sample Size Calculations

The following are two examples of formulas used to calculate appropriate sample sizes for your program evaluation. These are not the only formulas but they can be used in most situations that you find yourself. You may have another formula you prefer that is more than acceptable. These are here only as illustrative examples.

When calculating a sample size for any study be sure and keep in mind the level of analysis and what level of stratification you will ultimately be interested in looking at. If you want to compare males and females then you must have a sample size for each.

1. The following formula can be used to calculate change in key variables over time:

\[ n = D \left[ \sqrt{2P(1-P)} Z_{1-a} + \sqrt{P_1(1-P_1) + P_2(1-P_2)} Z_{1-\beta} \right]^2 / \Delta^2 \]

Where:
- \( D \) = design effect;
- \( P_1 \) = the estimated proportion at the time of the first survey;
- \( P_2 \) = the proportion at some future date such that the quantity \((P_2 - P_1)\) is the size of the magnitude of change it is desired to be able to detect;
- \( P = (P_1 + P_2) / 2; \)
- \( \Delta^2 = (P_2 - P_1)^2 \)
- \( Z_{1-a} \) = the z-score corresponding to the probability with which it is desired to be able to conclude that an observed change of size \((P_2 - P_1)\) would not have occurred by chance.
- \( Z_{1-b} \) = the z-score corresponding to the degree of confidence with which it is desired to be certain of detecting a change of size \((P_2 - P_1)\) if one actually occurred.

Example: For a one-sided test to measure change, then \( a=0.05 \) (\( Z_{1-a} = 1.65 \)) and \( b=0.20 \) (\( Z_{1-b} = 0.83 \))

<table>
<thead>
<tr>
<th>Power</th>
<th>1-sided</th>
<th>Alpha</th>
<th>2-sided</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>1.282</td>
<td>95%</td>
<td>1.645</td>
<td>95%</td>
</tr>
<tr>
<td>80%</td>
<td>0.83</td>
<td>90%</td>
<td>1.282</td>
<td>90%</td>
</tr>
<tr>
<td>70%</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Sample size calculation for a point estimate from cross-sectional survey (assumes a known standard error).

\[ n = (Z_{1-a} \times (p \times (1-p))) / Se^2 \]

\( Z_{1-a} \) = the z-score
\( p \) = the estimated proportion of the variable of interest
\( Se \) = known/acceptable standard error
10.3 Survey Respondent Consent Forms

This first example is a simple way to accomplish this part of the interview process. Some Ethical Review Boards give permission to use this shortened form. However, be aware that some Boards require much more extensive forms that include many more details (see the second example).

This first example incorporates the interview screening process with the consenting process.

**HIV/AIDS Behavioral Survey: Adults, 18 and older**  
**Country X**

**Screening question:**

<table>
<thead>
<tr>
<th>No.</th>
<th>Questions and filters</th>
<th>Response</th>
<th>Skip to</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>How old are you now?</td>
<td>Current age</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Estimate best answer</td>
<td>Must be at least 18 years old</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[<strong>/</strong>]</td>
<td></td>
</tr>
</tbody>
</table>

*Proceed with this questionnaire if the respondent is at least 18 years old. If the respondent is not yet 18 years of age, thank him or her for their time and move on to another potential respondent.*

**Introduction:**

My name is _____________ and I work with ________________. I am conducting important interviews with participants of our HIV program. You have been selected to participate in the interview. The purpose of this survey is to gather information on knowledge, attitudes, beliefs and behaviors (such as substance use and sexuality) related to HIV transmission among participants of our HIV programs and use that information to help us improve our program so that we can help people improve their health.

The answers you give here are completely anonymous; your answers will never be connected to you personally. I will not write your name on the questionnaire and there will be no way for anyone to know who answered these questions. Also, you cannot lose access to any services because of anything you say to me.

Since our program is related to HIV, some of the questions I am going to ask you are personal, including some questions about money, work and your personal sexual activity. This is NOT to judge you in any way; it is only to help us improve our programs. If at any time you do not want to answer any question, please tell me. You can also stop the interview at any time if you wish.

We will be asking these same questions of other adults. It is very important that you give me your most honest answers because this information will be used to improve the health programs that many people are involved in. If you do not understand a question, please ask me and I will try to explain it to you. If you agree, now I will begin.

**Consent:** May I ask you some questions?

| Yes...............1 (Begin the questionnaire) | No...............2  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondent signs or provides a thumb print</td>
<td>(Do NOT continue with the questionnaire)</td>
</tr>
</tbody>
</table>
This second example is much more extensive and provides not only more information about the study/evaluation but also contact information. The last page with the signature or thumb print is kept with the interviewer and the rest of the form is handed to the respondent once they have provided their consent to participate.

This process does not begin until after the potential respondent has been screened and found to meet all survey entry requirements.

**Name of Study**

INFORMED CONSENT FORM

**Introduction**

"Good morning/afternoon/evening. My name is ________. I am from an organization called (Name of Organization). We are conducting a study to learn about (provide your own justification for why this study is taking place). In order to be sure that you fully understand what it means to be in this study, I am going to read for you, or read with you, this consent form. I will also ask you to sign it, and I will give you a copy of this form. This consent form might contain some words that are unfamiliar to you. Please ask me to explain anything you may not understand.

This study is part of a project called (give name of project) that aims to (Restate the purpose of the study). It is funded by (provide name of funder).

**Reason for the Research**

You are asked to participate because you (give reason they are being asked to participate: they are the right sex, right age, live in the geographic area of the project... whatever the criteria for inclusion in the study are explain that to them). The Information will be collected (explain when and how often the data is being collected and how it relates to the project/program).

**Your Part in the Research**

You will be asked to answer a set of questions which I will ask you and I will record the answers you give me on this questionnaire. Mostly the questions will be about (list off the kinds of things in your survey, such as: the respondent, the respondent’s household; education and health of your children and other household members; including questions related to your household assets, income, and other health and social services that exist in your community that your household may or may not have access). All together you and I will need about (estimate how long this will take: 20 minutes, 45 minutes... 1 hour) to complete the questionnaire.

**Possible Risks and Benefits**

Your name will not be written down on the questionnaire. I will record your answers on this questionnaire, to enable me to correctly capture your answers correctly.

At any time during the interview, it is possible that some questions may make you uncomfortable. When this happens, you have the right to refuse to answer any question that you do not want to answer.
Your experience may help us *(tell them how this information once gathered will be useful for... Improving services, improving small business approaches, improve animal vaccination systems, etc...)*. However, you may not receive any direct benefit from participating.

**If You Decide Not to Be in the Research**

You are free to decide if you want to participate in the study. Your decision will not affect access to health and social services you would normally receive in your community.

**Leaving the Research**

If during the interview, you choose not to continue to participate, you are free to ask me to stop the interview at any time.

**Confidentiality**

This study is being carried out in a way that will keep your individual information confidential and no one will ever know your responses other than me as the interviewer. There may be some staff people who will look at the surveys to make sure they are properly completed and someone will be entering this data into a computer. Other than that your identity will always be completely confidential and will never be used. The information gathered from this study may appear in a paper, book, or other form for use by *(Give name of your organization)* and its implementing partner agencies. However, your name and identity will be kept private and there will be no information that links you to information you provided.

*(If using this for group discussions or Focus Group Discussions you can add this paragraph.)* We would like to remind you to keep the names of other participants and anything they say confidential, that means that none of us will talk about anything that we hear today once we leave this group. We will keep this discussion private. Please do not repeat anything you have heard today to anyone.

**Compensation**

You will not be paid to participate in the study.

Do you have any questions?

Please give them adequate time to formulate and articulate their questions.
If You Have a Problem or Have Other Questions
If you have a problem that you think might be related to taking part in this research or any questions about the study please call:

NAME OF LOCAL Principle Investigator,
Local Address,
City,
Local Telephone: #######
Email: person@email.org

If you have any questions about **your rights as a study participant**, call:

Name of local Ethics Committee Representative,
Name of Ethics Committee
Local Address,
City, Country
Local Telephone: ####-####-####-##
VOLUNTEER AGREEMENT

The above document describing the benefits, risks, and procedures for the study titled “Name of Study” has been read and explained to me. I have been given an opportunity to ask questions about the study and received satisfactory answers. I agree to participate as a volunteer.

____________________________________________
Printed Name of Volunteer

____________________________________________
Signature of Volunteer Date

OR If volunteers cannot sign the form, take their thumbprint.

____________________________________________
Printed Name of Volunteer

____________________________________________
Thumbprint of Volunteer Date

I certify that the nature and purpose, the potential benefits, and possible risks associated with participating in this research have been explained to the above individual.

____________________________________________
Printed Name of Person Who Obtained Consent (Research Assistant)

____________________________________________
Signature of Person Who Obtained Consent (Research Assistant) Date
10.4 Qualitative Methods Guidance

Focus Group Facilitation

Focus groups can help you understand why things are taking place in your community or uncover attitudes and norms that influence behaviors. You can also use FG to get a sense of what your respondents believe to be behaviors and patterns of behaviors in your community. One key question to ask that helps to find out about behaviors is to ask the group to talk about what they think that “others” (friends, people they know) are doing.

**Step one: Clearly define what the issues are that you need to find out about.**
Spend some time with colleagues deciding on what the general topics are that you want to find out about. Then decide if you want to find out about their: attitudes, beliefs, behaviors, norms, or values.

**Step two: Clearly segment out the populations you are interested in hearing from.**
Be very specific about who you want to have in your focus group and segment the population. Hold focus groups for each of these groups. Do not mix men with women and older people with younger people. There are dynamics between these various groups that may make some members of some groups go quiet and not talk. So have a FG with males between 15 and 19 and another one for males between 20 and 24, then a third one for females between 15 and 19 and a fourth one for females between 20 and 24. The groups should be as homogenous as possible.

**Step three: develop your focus group discussion guide**
Now that you know what you want to find out and who you want to ask, develop a guide that the facilitators will use to guide them through the discussion. This guide must cover all of the pertinent information that you want to find out about. You should develop it so that the questions you ask are open-ended; questions that cannot be answered with “yes” or “no” answers. Ask them in a way that leads the respondents to open up and talk.

**Step four: equipment you need**
To make sure that you gather all of the information your informants share, you will need to have a good tape recorder and a set of tapes to record the discussion. Afterwards you will probably want to have a good transcription Dictaphone to help with the transcription of the tapes.

**Step five: How to recruit respondents**
You can recruit respondents from within your target population by asking for volunteers from among your clients. You can do this either in the field or from among those who come to your site. It is best to run 5 or 6 groups for each segmented target group you have identified. However, since this gets expensive and time consuming you should hold at minimum 2 discussion groups for each target segmented target population. This will help to avoid having one discussion group with people who are at the extreme end of either end of the information spectrum.
Step six: Facilitating the discussion
The best focus group discussions are those that start with the introductory question posed by the group facilitator and then run on their own. These groups do not need the facilitator to intervene or ask any of the questions because the respondents are invigorated, interested and engaged. In the ideal setting, all members contribute and allow each other to share. In the real setting, the facilitator will need to pose the open-ended questions and listen. Then pose the “probe” questions that take the respondents further into that first question. These can be as generic as “Tell me more about that.” Or they may be specific probes that build off of the original question. “So you just said that young men are not ready to let go of their ‘sugar mommies’. Can you tell me how do these relationships start up?”

The facilitator guides the discussion by keeping an eye on the pre-determined questions on the guide and making sure that all questions are adequately discussed. Sometimes people will answer multiple questions at a time and jump from topic to topic. If the facilitator feels that the questions are spontaneously addressed they may not need to ask the question themselves. Also, it is acceptable for the respondents to answer the questions in any order, even if it is not the same order as the facilitator’s guide.

Always open the discussion by thanking the respondents from attending, ensure them that you will keep the information they provide confidential and that you will not be sharing this information with anyone outside of your study team. Then clearly explain what the topic is and begin the discussion. Close the discussion by asking if the respondents have any other thoughts or comments on anything that has been discussed that evening. Once they have fully spoken their minds thank them for their participation.

During these discussions the respondents sometimes ask for information from the facilitator or they may be giving incorrect information about the topic you are there to discuss. Keep both of these issues until after the discussion group is over before you address them. Simply tell the respondents that you can answer their technical questions about this topic at the end and that you can correct any misunderstood facts at that time as well; and continue with the discussion group.

Step seven: Analyzing the information
Once you have all the sessions transcribed the task becomes reading through the material for common themes and answers to your questions. Since this is a discussion, you will find responses to the question in multiple places. People will answer the questions as asked but will also go back to previously asked questions later on in the discussion. Entire books have been written on this so this document can not attempt to do this topic justice. Keep your field guide handy and read through the transcripts looking for and regrouping responses that address each question. Then synthesize those responses into a general description of the responses using exact quotes from respondents as needed to give examples of what was said.
Hearsay Ethnography

Hearsay ethnography is a method that is used to help get to what your target group is saying when you are not around. This method can help you in finding out how your population is relating to your intervention or how they are learning, not learning, understanding, not understanding or changing or not changing behaviors. This is ethnography so it does not allow you to generalize to populations beyond your own target group and should be used to give you a 'sense' of what is happening in your target population. It can be used to help you to find out whether or not your target group is making the changes in knowledge, attitudes, social norms and behaviors that you are hoping they are changing.

This method requires that you recruit several members of your target population “to write down, at the end of each day, what was said in conversations they overheard or participated in as they went about their normal activities. Although originally developed to augment data collected by demographers studying AIDS in sub-Saharan Africa, hearsay ethnography turns out to have broader methodological and theoretical advantages. Unlike surveys, focus groups, or in-depth interviews, hearsay ethnography captures spontaneous conversation, rather than responses to researchers’ questions or probes. Local residents can also gather data from many more locations and social contexts than a traditional ethnographer could do, including gender-segregated settings and those to which an outsider would not have access. More important, hearsay ethnography reveals the varied personal and collective agendas that lead people to confront AIDS, and it captures the ongoing shared enterprise of constituting and revising cultural understandings.

If they overheard anything concerning AIDS, they were to make mental notes of what people said and did, and then write their recollections in commonplace school notebooks that evening or soon thereafter.

They permit systematic analysis of a wide range of questions. Systematic analysis of the journals could, for example, trace changes in collective understandings as these evolve over time. Even if we do not know precisely what larger population our journalists represent, if over months or years the content of the conversations they overhear changes—if for example people begin to dismiss many negative images of condoms (Kaler 2004)—we may have some confidence that this represents a real change of opinion, rather than survey respondents’ increasing knowledge of what researchers want to hear. If over time conversations about AIDS become clearer and less confused—less likely to circle back on themselves or dissipate in uncertainty—that would indicate real growth in the collective sense of mastery of the issue.”

---

Key informant interviews

Key informants are experienced people with direct, expert knowledge of the subject you are studying. There are logical reasons why certain people should have expert knowledge in certain areas. Logically, some of the following should have expert knowledge of commercial sex, for example: experienced sex workers, especially informal leaders, experienced clients, management, security, reception and bar personnel in lodges and bars where sex work is common, doctors or nurses who frequently treat sex workers, traditional healers who provide services to sex workers, community based distributors of family planning services, social workers who work with sex workers, members of women’s groups and rape centers, police, though their comments should be interpreted cautiously, church workers, especially those with activities in areas where sex workers live, journalists who have reported on sex work, men in occupational categories where visits to prostitutes are frequent, including the uniformed services and transport workers; community leaders, especially in small, cohesive, established communities.

Within these groups, we ask people who the experts are. For example, when talking to sex workers, we ask them who their informal leaders or most experienced and knowledgeable members are. Similarly, we ask medical workers, journalists, social workers which of their colleagues work frequently with sex workers. People identified repeatedly as experts by different people may be promising key informants.

We usually talk repeatedly to key informants and once to other interviewees. However, we try to probe potential key informants to make sure they really are experts before we accept them as key informants and invest time interviewing them. When probing, we look for a long record of involvement, direct, personal experience and comments rich in situational and contextual detail and examples. We are wary of informants whose comments are limited to generalities. If our informant’s reply is satisfactory, we proceed; if not, we politely thank them and tactfully discontinue the conversation.

It is important to follow certain principles when doing key informant interviews. These include:

- Begin with truly exploratory, flexible, open-ended questions
- Never tell people they are wrong, give nonverbal clues or offer value judgments
- Share our own experiences (without disclosing strong views) if doing so relaxes informants
- Never move to a new topic until we have completely explored the topic under discussion
- Pursue all unanticipated, but important, issues that arise
- Make very brief notes on each informant’s comments
- Interpret and summarize the key informant interviews, perhaps using the following steps.

First, we make a list, partly for intervention purposes, of all the areas, categories of places, addresses, days and times where risky sexual activity occurs. Second, we produce a summary, in point form, of the key points made by each key informant. Third, we make a summary, also in point form, of the separate key informant summaries. We divide our
summary into areas of major and limited agreement and consider possible explanations for inconsistencies.

How do we know whether the information we get is reliable? There is no foolproof way, but we can do the following:

- We can be attentive to internal inconsistencies in the comments of key informants and explore these inconsistencies in a reassuring way.
- If some conclusions seem questionable, we try and determine whether an informant has drawn them from a single, memorable incident.
- We ask ourselves whether the key informant’s experience qualifies them to make a statement whose reliability seems uncertain.
- We consider carefully whether the attitude the informant holds towards sex work may have influenced particular answers.
- We compare answers of different key informants, looking for contradictions and points of consistency in the comments of different key informants.
- We compare the information gathered from key informants with that yielded by other methods.
## 10.5 Indicator Reference Sheet for the PMP

<table>
<thead>
<tr>
<th><strong>Descriptive characteristics of the indicator</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective</td>
<td>Which objective does this indicator lead to?</td>
</tr>
<tr>
<td>Name</td>
<td>The name should match what is in the PMP and the IPPT)</td>
</tr>
<tr>
<td>Definition</td>
<td>Provide a complete definition. Define terms in the definition if necessaryootnote{A lot of confusion arises because unknowingly people have different understanding of what an indicator is about. So give more details than less, including what the indicator is not if you think it may be useful}</td>
</tr>
<tr>
<td>Disaggregation</td>
<td>Indicate if the indicator has disaggregation categories</td>
</tr>
<tr>
<td>Units</td>
<td>This is not as straightforward as it looks. For example, for the units of an indicator on the level of investment or value of exports, should the numbers be in local currency or US $? In current or constant terms? If in constant terms, what is the reference year? And, is this a % or #? Is it cumulative?</td>
</tr>
<tr>
<td>Formula</td>
<td>When relevant, include the numerator and denominator if it is a percentage or a ratio, or the list of variables and formula if it is an index, etc. Note what the units are for the variables that make up the formula since this is data that will effectively be collected</td>
</tr>
<tr>
<td>Classification</td>
<td>Indicate if the indicator is to be reported cumulatively, as an increment between reporting periods, or as a level indicator. Cumulative means that the previous period value is included in the current period reporting value. Incremental means that only the value of the current period is reported, but the indicator could be reported as cumulative (it is recommended that indicators be reported as cumulative). Level means that the indicator value is time-bound and cannot be added or subtracted to the previous period value. Examples of level indicators are “Level of investment” and “Value of loans”</td>
</tr>
</tbody>
</table>

### Data information *(If a separate Data Flow Diagram exists, attach it. See section 8.2.1)*

<table>
<thead>
<tr>
<th>Source(s) of the data:</th>
<th>Such as government records, project records, survey, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods for collecting the data</td>
<td>Give as much details as possible. If it is a survey, is it a structured questionnaire with the data recorded on paper form or with PDA? Is the data self-reported or by direct observation? Who collects the raw data (give names and/or entity as relevant) and how is it stored before it is transmitted to Africare. Note the trail of the data until the indicator is finalized if a separate Data Flow Diagram is not attached.</td>
</tr>
<tr>
<td>Data collection frequency</td>
<td>Such as monthly, quarterly, annually, twice in the life of the project, etc. Various data used to construct an indicator may have different calendar. Also, the calendar for data collection is not necessarily the same as the reporting calendar. Provide a complete picture</td>
</tr>
</tbody>
</table>
### Data quality issues
Note when a data quality assessment has been done or is planned and what findings may have been found. Note any quality issues that need to be monitored.

### Baseline and Targets
Baseline and targets
Provide the date and value of the baseline and targets with details on their sources and construction. The assumptions and calculations used to set up the targets are important pieces of information that often get lost after the fact. Refer to another document if necessary.

### History and Narrative
Change to the indicator
Note any change that was made to the information above, with the date and reasons. Changes and adjustments are always necessary in the process of implementing and monitoring a project, but it should be documented. It is recommended to keep past versions of a reference sheet if the modifications to an indicator are significant rather than trying to detail everything in one current sheet.

Name and date
The name of the person responsible for the last update should appear, as well as the date.
### Illustrative Indicator Table for the PMP

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Baseline</th>
<th>Year 1 Target</th>
<th>Total Target</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project goal:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Objective 1:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR 1.1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR 1.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Objective 2:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR 2.1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR 2.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Objective 3:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Objective 4:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## 10.7 Advantages and disadvantages of various data collection methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desk review (also called literature review)</td>
<td>• There is almost always something available to get you started</td>
<td>• Assessing validity and reliability on secondary data is difficult.</td>
</tr>
<tr>
<td>Review of existing records and documents</td>
<td>• It is economical and efficient</td>
<td>• The information might be inaccurate.</td>
</tr>
<tr>
<td></td>
<td>• It is usually directly related to the culture and people of your program</td>
<td>• Sometimes there is no directly related information.</td>
</tr>
<tr>
<td></td>
<td>• Can help in determining value, interest, positions, political climate,</td>
<td>• Conducting a secondary analysis may be time-consuming.</td>
</tr>
<tr>
<td></td>
<td>public attitudes, historical trends or sequences</td>
<td>• There are sometimes ownership issues and sharing information is sometimes a</td>
</tr>
<tr>
<td></td>
<td>• Enables study of trends over time</td>
<td>challenge.</td>
</tr>
<tr>
<td></td>
<td>• It doesn't require an interviewing or intruding on people's lives</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>• This can be most useful when trying to monitor the quality of</td>
<td>• The potential for generalization is limited.</td>
</tr>
<tr>
<td>Informal/direct or in-depth observations</td>
<td>communication channels, such as peer educators</td>
<td>• In-depth observations are expensive and time-consuming.</td>
</tr>
<tr>
<td>over an extended period of time, e.g.</td>
<td>• Provides direct information about behavior of individuals and groups</td>
<td>• There is a need for well-qualified, highly trained observers.</td>
</tr>
<tr>
<td>inspection, field visits. Can be</td>
<td>• Allows for a good understanding of situation/context</td>
<td>• It is dependent on the observer's understanding and interpretation:</td>
</tr>
<tr>
<td>participatory or non-participatory</td>
<td>• Provides an opportunities for identifying unanticipated outcomes</td>
<td>selective perception of observer may distort data.</td>
</tr>
<tr>
<td></td>
<td>• Exists in natural, unstructured, and flexible setting</td>
<td>• It is dependent on level of trust between the observer and the observed.</td>
</tr>
<tr>
<td></td>
<td>• Enables observer to learn things that participants are unwilling to share</td>
<td>• The act of observing may affect the behaviors.</td>
</tr>
<tr>
<td></td>
<td>in interviews</td>
<td>• Investigator has little control over situation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Behavior or set of behaviors observed may be atypical.</td>
</tr>
<tr>
<td>Method</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Key informant interviews</td>
<td>- They usually yield richest data, details, new insights</td>
<td>- The time required to select and get commitment may be substantial</td>
</tr>
<tr>
<td></td>
<td>- They offer a flexible, in-depth approach</td>
<td>- The relationship between evaluator and informants may influence type of data obtained</td>
</tr>
<tr>
<td></td>
<td>- They are easy to implement</td>
<td>- Informants may interject own biases and impressions</td>
</tr>
<tr>
<td></td>
<td>- They provide opportunity to explore topics in-depth</td>
<td>- They may result in disagreements among individuals, leading to frustration/conflicts</td>
</tr>
<tr>
<td></td>
<td>- They provide information concerning causes, reasons, and/or best approaches from an &quot;insider&quot; point of view</td>
<td>- Their flexibility can result in inconsistencies across interviews</td>
</tr>
<tr>
<td></td>
<td>- Their advice/feedback increases credibility of study</td>
<td>- The volume of information can be too large; may be difficult to transcribe and reduce data</td>
</tr>
<tr>
<td></td>
<td>- They allow interviewer to explain or help clarify questions, increasing the likelihood of useful responses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- They provide a pipeline to pivotal groups</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- They may help to solidify relationships between evaluators, clients, participants, and other stakeholders</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Their advice/feedback increases credibility of study</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- They allow assessment of agreement/disagreement on a given issue</td>
<td></td>
</tr>
<tr>
<td>Focus group interviews</td>
<td>- They are of reasonable cost and efficient</td>
<td>- There is a risk of one-sidedness on the part of participants and the moderator</td>
</tr>
<tr>
<td></td>
<td>- They help generate ideas</td>
<td>- Only a limited number of questions may be asked</td>
</tr>
<tr>
<td></td>
<td>- They usually yield richest data, details, new insights</td>
<td>- The volume of information can be too large; may be difficult to transcribe and reduce data</td>
</tr>
<tr>
<td></td>
<td>- They permit face-to-face contact with respondents</td>
<td>- They may reinforce existing power relation in the groups</td>
</tr>
<tr>
<td></td>
<td>- They provide opportunity to explore topics in-depth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- They provide quality control through people’s interaction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- They allow people’s views to develop in exchange with others</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- They allow assessment of agreement/disagreement on a given issue</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Formal survey | • Is most suited to answering questions such as what, how many, and how often  
• They permit face-to-face contact with respondents  
• Produces statistically reliable information  
• Results may be generalized if informants belong to a representative sample  
• Results may be perceived as the most reliable and objective by external observers  

Oral interviews or written questionnaires in a representative sample of respondents | • Data collection is demanding  
• They are intrusive  
• It is expensive  
• Data collection and analysis demand time  
• They require well-qualified, highly trained interviewers  
• Interviewees may distort information through recall error, selective perceptions, and/or a desire to please interviewer  
• It is inflexible: duration, questions, sample must be defined at the inception and cannot be changed |
| Informal survey | • Is reasonable in scope and resources needed  
• Not as costly as formal surveys  
• They are usually carried out rapidly  
• Provide results in a short period of time  
• They permit face-to-face contact with respondents  
• They provide opportunity to explore topics in-depth  

Involves quantitative surveys of small samples | • There is risk of sampling errors/biases  
• Is less suited for generalization |
| RAP, RRA, PRA | • Is fast  
• Is flexible  
• Can provide a qualitative, in-depth view on specific problems  
• Builds buy-in and ownership  

A participatory methodology comprised of a range of methods and tools. Typically uses key informant interviews, group interviews, community meetings or dialogue processes, as well as observation combined with techniques such as mapping, diagramming, seasonal calendars, ranking and the transect walk | • Generalization is limited  
• Bias may creep into the results  
• Relies on interpretation  
• Data collectors and analysts are one and the same; requires highly skilled data collection  
• Badly managed group exercises can take too much people’s time, ignore some people’s view, and legitimate the views of the dominant groups |
10.8 Evaluation Research Protocol Template

AFRICARE
RESEARCH PROPOSAL

TITLE:

PARTNER ORGANIZATIONS: Implementation/research TBD

PRINCIPAL INVESTIGATOR: TBD
PRINCIPAL PARTNER CONTACT STAFF: TBD
TBD

DURATION: Start date:
Proposed start/end dates, total months End date:
Duration:

DRAFT DATE:
I. Background

II. Study Objectives

III. Methodology
1. Design
2. Study sites
3. Study procedures
4. Key indicators
5. Study instruments

IV. Study population
1. Population
2. Inclusion criteria
3. Exclusion criteria
4. Sample size calculation
5. Sample frame
6. Recruitment

V. Statistical analysis

VI. Protection of Human Subjects
1. Consent process
2. Potential risks
3. Potential benefits
4. Ethical Clearance

VII. Dissemination of results

VIII. Timetable
IX. References

X. Annex
Voluntary Informed Consent Forms

Data collection instruments