

MODULE IV

MONITORING AND EVALUATION

PURPOSE

With the completion of Modules I, II and III of *ProPAN*, the team has outlined the main infant and young child feeding problems and main strategies through which these can be improved. During the course of the program implementation, how will the team know if the program is being implemented as it was designed? How will they know if improvements need to be made to its design? How will they know if the program actually improved child feeding problems? These questions can be answered with a well-designed *monitoring* and *evaluation* plan. This module provides information relevant to the development of such a plan. Additional resources may need to be consulted, depending on the scope of the plan and experience of those implementing it.

In practice, behavior-change programs have sometimes failed to significantly improve child nutritional status. The main reasons that programs fail can be grouped into three categories:

- ◆ The program was poorly implemented.
- ◆ The program designers made a poor choice of which behaviors to promote.
- ◆ Although well implemented, the program did not have the effects on behaviors that were anticipated.

For example, an educational program designed to promote the consumption of carrots by children might not have the expected impact on vitamin A status because the educators did not properly teach mothers¹ the importance of giving carrots to children. Or else, it could be that mothers¹ learned about the importance of giving carrots, but were not able to purchase them during certain months of the year. Another reason would be that, although children were given carrots more frequently than before the program started, the amounts given were still insufficient to increase vitamin A status to adequate levels.

¹ The vast majority of young children are likely to be cared by their mothers. However, we used “mother” throughout *ProPAN* to denote mothers and other caregivers.

Therefore, the best way to assure that the program designed is being properly implemented and has the intended effects is through adequate attention to monitoring and evaluation. The objective of this module is to create an integrated monitoring and evaluation plan to:

- ◆ Assess the progress of the program designed in Module III (monitoring).
- ◆ Determine if any modifications should be made to the program design (monitoring).
- ◆ Evaluate the outcomes of the program designed in Module III (evaluation).

PRODUCT

Upon conclusion of this module, the team will have a monitoring and evaluation plan that describes how the program designed in Module III will be periodically assessed to determine if it is proceeding according to plan and evaluated to determine if it has had the expected outcome on infants and young children.

STEPS

The conceptual framework for this module stems from a World Bank publication on monitoring and evaluating nutrition programs (Levinson *et al.*, 2000). Per this publication, the steps in creating a monitoring and evaluation plan are as follows:

- ◆ Specification of program goals.
- ◆ Identification of program inputs, outputs, outcomes, impacts and benefits.
- ◆ Design of a monitoring and evaluation system.

Each one of these steps is discussed below.

DEVELOPMENT

Step 1. Specification of program goals

In Module III, a series of intervention strategies were developed. In the first step of Module IV, these strategies are revisited to articulate the overarching goals of the program.

Levinson *et al.* (2000) offer the following definition for goals:

“Goals are the broad aims of the program, the significant, longer-term changes that planners expect to occur in people’s lives. For example, the reduction of severe protein-energy malnutrition and the significant reduction of iodine deficiency disorders”.

In other words, goals are focused on improvements in the nutritional status of young children. As such, a program focused on young child malnutrition might have goals like reducing malnutrition and reducing vitamin A deficiency.

Step 2. Identification of program inputs, outputs, outcomes, impacts and benefits

Levinson *et al.* (2000) describe a five-component framework that makes explicit how program activities will help programs meet the goals articulated in Step 1 (Table IV-1, Form IV-1).

Table IV-1. Monitoring and evaluation framework that details how program activities will lead to achievement of program goals (Levinson *et al.*, 2000).

Goals: Broad aims of program				
Inputs →	Outputs →	Outcomes →	Impacts →	Benefits
Resources used to support the primary activities of the program.	The delivery of goods and services.	Changes in behaviors or practices.	Nutritional status measures.	Broader effects.
<i>Monitoring</i>		<i>Evaluation</i>		

These five components, expressed quantitatively, are defined as follows:

Input are the materials, goods and services needed to implement the program. Inputs include such items as training of program personnel, educational materials, food, supplements, equipment to measure children’s weight and height, and growth charts

Outputs refer to the successful delivery of the program’s materials, goods and services to the target population. For example, in a communication program, an output would be the number of mothers that received education on incorporating iron-rich food into their children’s diet. Other outputs include the number of program personnel trained on a particular topic, educational materials delivered to the intended population, food provided to target families, supplements given to children, equipment delivered to health personnel or facilities, and growth charts given to children’s families.

Outcomes those effects the program outputs can have which are necessary for the program to have the intended impact. Change in feeding behavior is an example of an outcome. For instance, caregivers feeding children more iron-rich foods is an intended result of a program, but does not guarantee that the program has had an impact on children’s nutritional status.

Impacts are the biological and/or observable changes in the nutritional status of the child. Anthropometric (e.g. growth) and micronutrient measures are commonly used measures of the impact of a nutrition program. Impacts could include reduction in the prevalence of anemia, incidence of stunting, or incidence of low birth weight.

Benefits are the ultimate effects that the program hopes to have, but are unlikely to be measurable within the context of an on-going program. Benefits of a successful nutrition program may include such effects as increased cognitive ability, decreased morbidity, increased work productivity, increased life expectancy, and decreased mortality.

Through monitoring, the quantity and quality of inputs and *outputs* are documented. Through evaluation, *outcomes*, *impacts* and *benefits* are measured.

The decision to measure impacts and benefits will depend on the specific questions that need to be answered as well as the evaluation budget available. For some programs it may not be feasible or advisable to measure impacts and benefits. It is important to note that both weight and length are affected by a wide range of factors in addition to diet and feeding practices. Length is also difficult to measure in the field and is only likely to be affected when the intervention occurs prior to the second year of life. When resources are available, weight and length should be included in program evaluations², however, even well run programs that positively affect feeding practices may not at the same time demonstrate significant changes in length.

As mentioned, benefits are longer-term effects that require a long follow-up period and substantial resources to evaluate properly. For these reasons, benefits are rarely evaluated in programs. Thus, most programs evaluate their effect on outcomes, in other words, the specific behaviors being promoted. Often, impacts are also evaluated.

Table IV-2 presents an example of these monitoring and evaluation components. A program in which the goal was to reduce anemia among Peruvian women and children, was developed whereby cooks from “community kitchens” (*comedores populares*) were trained (inputs) to increase the use of chicken livers in recipes (outputs) (Creed-Kanashiro *et al.*, 1998). It was expected that with the increased availability of chicken liver in the recipes (outputs), women and children attending the “community kitchens” would consume more chicken livers in their diet (outcomes). The additional iron in their diet, from the iron-rich chicken liver source, would increase their circulating blood iron (hemoglobin) and decrease anemia among women and children (impact). With adequate iron status, children would have improved learning capacity and women would have increased work capacity (benefits).

2 Guidelines on how conduct anthropometric measurements can be found in Cogill B (2003) Anthropometric Indicators Measurement Guide. Washington, DC: Food and Nutrition Technical Assistance Project, Academy for Educational Development. (Available online at www.fantaproject.org.)

Table IV-2. The monitoring and evaluation conceptual framework: An example from a Peruvian program (Creed-Kanashiro *et al.*, 1998).

Goals: Reduction of anemia among women and children				
Inputs →	Outputs →	Outcomes →	Impacts →	Benefits
Training of “community kitchen” cooks Adequate supplies of chicken livers	Chicken livers provided three times weekly to women and children attending “community kitchens”	Increased consumption of chicken liver at “community kitchens” by women and children	Reduced prevalence of iron-deficiency anemia among women and children	Improved learning and work capacity
<i>Monitoring</i>		<i>Evaluation</i>		

In Module III, several intervention strategies were discussed. As shown in Table IV-3, different intervention strategies can have different inputs, outputs and outcomes. Often, however, the different inputs will converge on similar outputs and outcomes.



Table IV-3. Sample inputs, outputs and outcomes from intervention strategies described in Module III.

Intervention Strategy	Sample Inputs	Sample Outputs	Sample Outcomes
Training	Health volunteer training to demonstrate recipes using iron-rich food to mothers	Targeted mothers receive recipe demonstration	<ul style="list-style-type: none"> • Targeted mothers make iron-rich recipes • Targeted children consume iron-rich meals • Targeted children increase their iron intake
Norms (to change hospital norms so that all newborns are put immediately to the breast)	Meetings arranged with hospital staff in charge of maternity ward norms	Meetings held with staff in targeted hospitals	<ul style="list-style-type: none"> • Targeted hospitals change protocols so that all newborns are put immediately to the breast • Targeted hospitals implement new protocols
Communication	Development and production of IEC materials on increasing feeding frequency	Targeted parents receive materials	<ul style="list-style-type: none"> • Targeted children’s feeding frequency increases • Targeted children increase their caloric intake
Community participation (to organize breastfeeding support groups)	Meeting space identified for support groups	Meeting space provided for breastfeeding support groups	<ul style="list-style-type: none"> • Targeted mothers exclusively breastfeed for 6 months
Coordination with strategic allies	Agreement with poultry processing plants to provide chicken livers at cost	Chicken livers are provided at cost	<ul style="list-style-type: none"> • Targeted children increase their consumption of chicken livers
Advocacy (to enlist pediatricians’ associations in the promotion of the introduction of complementary foods at 6 months of age)	Meetings arranged with cognizant officers in pediatrician associations	Meetings held with officers in targeted associations	<ul style="list-style-type: none"> • Targeted associations that introduce bylaws to promote the introduction of complementary foods at 6 months of age
Legislation (to fund growth monitoring equipment and activities in rural areas)	Development of draft legislation	Draft legislation presented to legislators representing rural constituents	<ul style="list-style-type: none"> • Rural children increase attendance at growth monitoring and promotion activities
Production of fortified foods	Agreement with food company to develop a fortified complementary food	Fortified complementary food produced and distributed	<ul style="list-style-type: none"> • Targeted children increase their consumption of fortified complementary food

Step 3. Design of a monitoring and evaluation system

3.1. MONITORING SYSTEM

Levinson *et al.* (2000) recommend considering what role program personnel can play in monitoring and evaluation. Is the required information to gather so specialized or time consuming that program personnel will be overburdened? Should an external institution be invited to gather all or some of the monitoring and evaluation data? Depending on the size and scope of the program, it is likely that employing a mix of both internal and external personnel is best. External personnel are likely to be more objective in their assessment and may have some specialized knowledge (for example, in statistics), which can be useful for a more rigorous evaluation. Internal, program personnel, on the other hand, will certainly be more familiar with the history and setting of the program and will be able to provide important contextual insights. Use of internal program personnel also allows for some capacity building for the future.

Assuming the team has primary responsibility for monitoring and evaluation, it must carefully consider which information is required for decision making and how often it needs to be updated. Factors to assess are:

- ◆ What information will be most useful for monitoring the program? (For example, will knowing the religious affiliation of the families of children who attend growth monitoring sessions help the team to improve the program as much as knowing the children's age?)
- ◆ What information is already being routinely collected? (For example, if a program is implemented within an existing larger program or at a health facility, it is likely that information on growth monitoring, clinic attendance, supplement distribution is being collected.)
- ◆ What information can be feasibly collected on an on-going basis by program staff? (For example, if someone spends 15 of 20 working days a month recruiting, training, and supervising health volunteers, can this person reasonably be expected to sort through clinic records for information on 50-100 children per month?).

Although outcomes fall in the purview of evaluation in the conceptual framework presented in Table IV-1, information on outcomes should be collected regularly during program implementation and not just at evaluations. The mechanism for gathering this information on an ongoing basis is the monitoring system.

A framework is provided to aid with assessing these three factors (Table IV-4, Form IV-2). Input, output and outcome information to be gathered and the frequency of data collection are noted in the first two columns. Then, these questions are evaluated one by one: (1) Is this information useful? (2) Is it already collected? (3) Can it be feasibly collected? After going through this process, the team will decide on the amount and kind of information to gather for program monitoring. See Table IV-5 for an example.

Table IV-4. A framework for deciding what monitoring information needs to be collected and how frequently

Information	Frequency	(1) Useful?	(2) Already collected?	(3) Feasibly collected?
Input				
Output				
Outcome				

Table IV-5. An example of a completed framework for deciding what monitoring information needs to be collected and how frequently

Information	Frequency	(1) Useful?	(2) Already collected?	(3) Feasibly collected?
Input: Training of community cooks				
Number of trainings held	1 per month	Yes	Yes	Yes
Number of trainees who attended	1 per month	Yes	Yes	Yes
Number of trainees who passed a post-test assessing minimum skills and knowledge	1 per training	Yes	No	Yes
Age of trainees	1 per training	No	No	Yes
Sex of trainees	1 per training	No	No	Yes
Education level of trainees	1 per training	No	No	Yes
Ethnicity of trainees	1 per training	No	No	Yes

Although all of the input information can be feasibly collected, the team decides that the first three are the ones which will yield the most useful information. Forms will have to be developed to assess “Number of trainees who passed a post-test assessing minimum skills and knowledge” because this information is not being collected through an existing data-collection or monitoring system.

3.2 . INDICATORS

How can the monitoring and evaluation information gathered be most helpful to the program manager and implementing team? By being carefully crafted to measure what is intended. This is achieved through the development of indicators.

Levinson *et al.* (2000) offer the following as characteristics of good indicators: simple, clearly defined, measurable, variable, valid, reliable and quantifiable. They emphasize that useful indicators “must show variation between persons and over time” or else they “will not discriminate between those who have benefited from the program and those who have not.” For example, child height varies over time and between well and poorly nourished children whereas adult height does not, making height a good indicator for children’s, but not for adults’, nutritional status.

Next, forms need to be developed for all of the monitoring information that will be collected, field personnel need to be identified, and the frequency of data collection needs to be specified. Form IV-3 provides a matrix to aid with the planning of data collection and an example of a completed matrix is shown in Table IV-6 below.

Table IV-6. Example of a matrix for monitoring and evaluating program activities
(adapted from Levinson *et al.*, 2000)

Indicator	Form	Person	Frequency
Input: = Number of community cooks who passed post-test Number of community cooks who attended training	Training form	Nutritionist field worker	<ul style="list-style-type: none"> • Monthly
Output: = Number of days per week chicken livers were incorporated into recipes	Menu review form	Nutritionist field worker	<ul style="list-style-type: none"> • Monthly
Outcome: = Number of parents reporting that their children ate chicken liver at community kitchens Total number of parents interviewed with children less than 2 years	Food frequency questionnaire	Nutritionist field worker	<ul style="list-style-type: none"> • Every 3 months • Baseline • Endline
Impact: = Number of children with hemoglobin values below iron-deficiency cutoff Total number of targeted children	Blood indexes form	Laboratory technician	<ul style="list-style-type: none"> • Baseline • Endline

In this example, some of the indicators are assessed with a single number (for example, the output indicator) and others with a ratio of two numbers (for example, the input, outcome and impact indicators). The “form” column indicates the name of a form that needs to be developed to gather information on that specific indicator. The “person” column refers to the technical background of the person or group of persons who will be responsible for filling out the form. Finally, the frequency with which monitoring information is going to be collected and reviewed and when the information will be collected for evaluation purposes are noted in the “frequency” column. In the case of the input indicator, even though information will be collected at each training (see Table IV-5) and several training workshops are held each month, the information will only be synthesized into the indicator once per month.

3.3 Evaluation system

Evaluation of the outcomes and impacts of a program, i.e., whether the program has had the anticipated effects and whether the observed effects, if any, were due to the program, requires the following:

- ◆ Choosing the evaluation design.
- ◆ Determining the sample size and sample frame.
- ◆ Identifying the control group.

These will be discussed in turn.

3.3.1. CHOOSING THE EVALUATION DESIGN

There are many evaluation designs from which to choose, a fact that frustrates and confuses many program planners. The most rigorous design is a **randomized controlled trial** in which a program is randomly allocated to one group of individuals (or communities) while a comparable group (the control group) receives no program. While the proper implementation of this approach provides the best evidence of the effect (or non-effect) of a program, it requires considerable resources.

There are a number of alternative evaluation designs that are less rigorous than the randomized, controlled trial (Fitz-Gibbon, Morris, 1987; Levinson *et al.*, 2000). Regardless of the design, all evaluations should include *at least a baseline* and an *endline* evaluation. Many program managers make the mistake of not conducting a baseline survey prior to the initiation of program activities. Comparing outcome and impact indicators in the program group to the same indicators in a control group only at the end of the program is a commonly used but very weak design. For comparison, several evaluation designs are summarized in Table IV-7.

Table IV-7. Types of program evaluation designs (adapted from Levinson *et al.*, 2000).

Design	Comments
(1) Pre-post design with a randomized control group	The most rigorous of designs. Often not feasible due to lack of resources or other constraints.
(2) Pre-post design with a non-randomized control group	Differs from (1) only in that the control group is not randomly assigned. The control group is purposively selected to be as similar as possible to the intervention group. Clearly, extreme care should be taken in selecting this control group. It is recommended that statistical or epidemiological assistance be obtained before making a final decision.
(3) Pre-post design without a control group	The lack of a control group identified at the start of the program makes it much more difficult to attribute changes seen in the intervention group to the program. Programs that use this approach should attempt to gather information on other factors that might be causing changes in behaviors and/or nutritional status (for example, government radio spots or changes in food prices or availability).
(4) Time-series design	This requires multiple measures of indicators <i>before</i> the start of the program to establish a trend in the direction and magnitude of change of the indicator (for example, the proportion of children < -2 weight-for-age Z score). The analysis then examines whether the rate of change was significantly hastened after the initiation of the program.
(5) Institutional-cycle design	Many nutrition programs are “phased in” over a number of months or years due to limitations in personnel or resources. This rolling implementation can be used to the advantage of the program planner by taking a series of before-implementation measurements at each successive round, making it comparable to design option (1) or (2) depending on whether the phase-in is done randomly or not.

3.3.2. DETERMINING THE SAMPLE SIZE AND SAMPLING FRAME

Once an evaluation design has been selected, the next steps are to determine the sample size and choose the sampling frame (Levinson *et al.*, 2000). Knowledge of how to determine the sample size and choose the sampling frame is quite specialized so it is recommended that program planners enlist a statistician or epidemiologist for assistance. In preparation for the discussion of these issues with an expert, these items should be clearly defined (Levinson *et al.*, 2000):

- ◆ The number of groups of interest being studied, if more than one (for example, rural and urban participants, communities receiving all program activities and those receiving a subset of activities).
- ◆ The key indicators that will describe outcomes (for example, the proportion of children consuming meat products daily).
- ◆ The amount of change expected in each indicator (for example, increase in the proportion of children consuming meat products daily from 30% to 40%).

With regards to the “amount of change expected in each indicator,” programmers tend to hypothesize that their programs will have effects that are *larger* than is realistically achievable. For instance, a 25% decrease in malnutrition (for example, in height-for-age Z score), would be very difficult to achieve through a single program implemented over a relatively short 3 to 5-year time frame.

The statistician will use these three pieces of information, along with a measure of the level of confidence needed to draw conclusions and the probability of detecting a difference in the indicator, if one actually exists, to calculate the **sample size** needed to test the hypothesized effects.

The other specialized step is the selection of the **sampling frame**. Options include simple random sampling, stratified sampling, systematic sampling, and cluster sampling (Levinson *et al.*, 2000). The choice of which of these sampling frames to use depends to a great extent on such factors as how dispersed the population being studied is and the existence of a complete population registry. Information on these factors will facilitate the discussions with the statistician.

3.3.3. IDENTIFYING THE CONTROL GROUP

In identifying a control group, it is important to keep in mind that “individuals (or communities) in a control group must be identical, or as similar as possible, to the group of program participants (Levinson *et al.*, 2000).” The best way to achieve this is to randomly select the communities or individuals that will participate in the program and those that will not. If the program will be phased in over time to different communities, those that will receive the program activities first can be randomly selected while those receiving the program last can be assigned to be the control group.

3.4 Developing a timeline

Finally, all monitoring and evaluation activities should be specified in a timeline that considers program implementation activities outlined in the intervention plan (Module III) as well. See Table IV-8 for an example.

Table IV-8. Example of a timeline of program activities, including monitoring and evaluation.

Activity	Year 1				Year 2			
	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Complete intervention plan and monitoring and evaluation plan	X							
Develop forms for gathering monitoring and evaluation data	X							
Hire and train field workers for baseline evaluation	X							
Carry out and analyze baseline evaluation	X							
Develop curriculum for training community cooks		X						
Train community cooks in adding chicken liver to recipes		X	X	X	X	X		
Collect monitoring information on training of community cooks		X	X	X	X	X		
Collect monitoring information on chicken livers incorporated into recipes by community cooks		X	X	X	X	X	X	X
Carry out and analyze endline evaluation								X

REFERENCES

- Creed-Kanashiro H, Uribe T, Bartolini R, Fukumoto M, Villasante R, Zavaleta N, Bentley M. (1998) Intervención Educativa para Mejorar el Consumo de Alimentos Ricos en Hierro y Prevenir la Anemia en Mujeres y Niñas Adolescentes a través de los Comedores Populares. Lima, Peru: Instituto de Investigación Nutricional.
- Fitz-Gibbon CT, Morris LL (1987) How to Design a Program Evaluation. Newbury Park: Sage Publications.
- Levinson FJ, Rogers BL, Hicks KM, Schaezel T, Troy L, Young C (2000) Monitoring and Evaluation: A Guidebook for Nutrition Projects Managers in Developing Countries. Boston: International Food and Nutrition Center.