Photo Credit: USAID/Nepal

BUREAU HUMANITARIAN ASSISTANCE

GUIDANCE FOR MONITORING, EVALUATION, AND REPORTING FOR EMERGENCY ACTIVITIES

December 2020

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

AAP	Accountability to Affected Populations
ADS	USAID's Automated Directives System
AOR	Agreement Officer's Representative
APS	Annual Program Statement
AR	Annual Report
ART	Awards Results Tracking System
BBS	Beneficiary based survey
BHA	USAID Bureau for Humanitarian Assistance
DDL	USAID Development Data Library
DEC	USAID Development Experience Clearinghouse
DQA	Data Quality Assessment
DRR	Disaster Risk Reduction
FCS	Food Consumption Score
FEWS NET	Famine Early Warning Systems Network
FPR	Final performance report
FY	Fiscal year
HH	Households
HHS	Household Hunger Scale
IDP	Internally displaced person
ITT	Indicator Tracking Table
IYCF	Infant and young child feeding
LOA	Life of award
LQAS	Lot quality assurance sampling
LRIP	Local, regional, and international procurement
M&E	Monitoring and evaluation
MPCA	Multipurpose cash assistance
MT	Metric ton
NFI	Non-food items
NGO	Non-Government Organization
NPE	Non-permissive environment
PBS	Population based survey
PDM	Post-distribution monitoring
PIO	Public international organization
PPS	Probability proportional to size
PVO	Private voluntary organization
PIRS	Performance Indicator Reference Sheet
R	Required indicator
rCSI	Reduced Coping Strategies Index
RiA	Required if applicable indicator
RTE	Real-time evaluation
SAR	Semi-annual report
SOW	Statement of work
SRS	Simple random sampling
TPM	Third-party monitoring
UN	United Nations
USAID	U.S. Agency for International Development
USG	United States Government

CHAPTER 1: SUMMARY OF MONITORING, EVALUATION, AND REPORTING PROCESSES

1.1 BACKGROUND

The mission of the U.S. Agency for International Development's (USAID) Bureau for Humanitarian Assistance (BHA) is to partner with other actors to provide international humanitarian assistance, alleviate suffering, and promote human welfare to the world's most vulnerable populations. Through its emergency awards, BHA provides life-saving humanitarian assistance and disaster risk reduction (DRR) that reduces suffering, and supports the early recovery of populations affected by both acute and protracted emergencies. BHA responds to emergency situations, or complex crises, and seeks to help internally displaced people who have been forced to flee their homes, as well as providing food assistance to refugees who have crossed national borders.

The primary purposes of monitoring, evaluation and reporting for BHA emergency activities are to:

- Fulfill BHA's obligation to ensure the effective and efficient use of resources; and
- To support adaptive management decisions to achieve the best possible outcomes for beneficiaries.

This document describes key monitoring, evaluation, and reporting responsibilities of BHA international emergency assistance awards using Title II or International Disaster Assistance (IDA) accounts. The guidance applies to activities implemented by U.S. or non-U.S. non-governmental organizations (NGOs), including private voluntary organizations (PVOs). The guidance outlined in this document does not apply to public international organizations (PIOs), although they are encouraged to use this document as a resource. This guide is intended to provide supplementary technical guidance to what is communicated through the BHA Emergency Application Guidelines and award language. For further information regarding application submission and award process for grants and cooperative agreements, please refer to the BHA Emergency Application Guidelines 2021-2022.

Key Terms and Definitions

In order to achieve a common understanding of terminology, definitions, and their appropriate use, the following terms have been defined per USAID's ADS Chapter 201, Program Cycle Operational Policy, as follows:

Monitoring: The ongoing and systematic tracking of data or information relevant to USAID's policies, operations, programs, Strategies, projects, and activities. Relevant data and informational needs are identified during planning and design, and can include output and outcome measures directly attributable to or affected by USAID-funded interventions, as well as measures of the operating context and programmatic assumptions. Monitoring informs strategy, project, and activity design and implementation. The analysis of monitoring data should inform progress towards anticipated results, efforts to manage adaptively, and promote accountability.

Evaluation: The systematic collection and analysis of data and information about the characteristics and outcomes of one or more organizations, policies, programs, strategies, projects, and/or activities conducted as a basis for judgments to understand and improve effectiveness and efficiency, timed to inform decisions about current and future programming. Evaluation is distinct from assessment (which is forward-looking) or an informal review of projects.The purpose of evaluations is twofold: to ensure accountability to stakeholders and to improve design, implementation, and BHA policy and guidance.

For the purposes of this document, **reporting** refers to the semi-annual, annual, and final reporting processes that provide updates on the programmatic progress and compliance of BHA emergency awards. Reporting requirements are stipulated in the terms of the award, which may reference other documents such as the BHA Emergency Application Guidelines for FY21 and FY22, and the Annual Report (AR) guidance.

Partners must consider the principle of **Do No Harm** when designing M&E systems, paying attention to who is collecting data, from whom, where, when, and how. This is important to consider when collecting sensitive information. It is also an important consideration when implementing in conflict-affected areas and/or in the context of a pandemic, where partners must balance the tradeoffs between collecting enough data to verify their activities with the potential security or health risks facing their staff and beneficiaries.

Box 1. USAID and BHA Terminology

*Note that the USAID's Automated Directive System Chapter 201 defines how the terms program, project, and activity should be used. In short, a **program** includes projects and activities that are aligned with a USAID Mission Country Development Cooperation Strategy (CDCS) Development Objective. **Projects** are groups of activities or other awards that are designed to achieve intermediate results within a USAID Mission CDCS or USAID Bureau results framework. Projects, in other words, are made up of individual activities or awards, which are implemented by partners, e.g., private voluntary organizations. **Activities** are awarded to partners using grants, cooperative agreements, bilateral agreements, contracts or other mechanisms. Each activity carries out an **intervention** or set of interventions. Implementing agencies will apply for activities under the BHA application guideline to carry out a set of interventions for emergency response (e.g., delivery of food assistance, training of community health workers). Each activity should have a clearly articulated theory of change and indicator tracking table, as outlined in the BHA Guidelines, which articulates how **output** and **outcome** indicators will be used to track performance toward an activity's stated Goal, Purpose(s), Sub-Purpose(s)*, Intermediate Outcome(s)*, Outcomes, and Outputs. (*Note that these components are optional.)*

1.2 MONITORING, EVALUATION AND REPORTING REQUIREMENTS

Monitoring and evaluation requirements for BHA emergency awards vary by award length, as shown in Table 1 below. Awards of six months or longer in duration are required to conduct a baseline and endline study. If the length of the award is 18 months or longer, partners are required to conduct an evaluation. The evaluation requirement also applies if your organization has implemented at least one BHA-funded award (of any duration, in any sector) in the past three years in a given country and your

organization has not completed an evaluation of any BHA-funded awards in that given country in the past three years. Partners must complete at least one evaluation of any BHA-funded award(s) at least once every three years in a given country. Exceptions to that requirement are listed in Table 1 below. If an applicant plans to use an exception, the justification should be included in its M&E Plan at application.

Table 1: M&E Requirements by Award Length

AWARD LENGTH	M&E REQUIREMENTS	
	APPLICATION	**POST-AWARD**
Less than six months	• Indicator Tracking Table (ITT) • M&E Plan Narrative	• ITT updated with baseline values and PIRS for custom indicators (Due within 90 calendar days of award) • Endline indicator values submitted with Final Performance Report
Six months or longer	• Indicator Tracking Table (ITT) • M&E Plan Narrative ○ Abbreviated baseline/endline SOW integrated in Monitoring Approach Narrative	• ITT updated with baseline values and PIRS for custom indicators (Due within 90 calendar days of award) • Baseline Report (Due within 90 calendar days of award) • Endline indicator values submitted with Final Performance Report

18 months or longer	• Indicator Tracking Table (ITT)
	• M&E Plan Narrative
	  ○ Abbreviated baseline/endline SOW integrated in Monitoring Approach Narrative
	  ○ Abbreviated Evaluation SOW integrated in Evaluation Approach Narrative
		• ITT updated with baseline values and PIRS for custom indicators (Due within 90 calendar days of award)
		• Baseline Report (Due within 90 days of award)
		• Full Evaluation SOW (Due 6 months prior to start of evaluation)
		• Evaluation Report (Due within 90 calendar days of end of activity)

Exceptions to the requirements can be requested for: • Responses immediately following a sudden-onset disaster (e.g. hurricane, earthquake, tsunami, flood, cyclone). • If a cost modification extends the length of the award to 18 months or longer, the partners should propose whether adding a final evaluation is appropriate in the modification application.		

Note: BHA reserves the right to require an evaluation of the proposed activity even if it does not meet one of the above criteria.

In addition to the requirements outlined in Table1, and the remainder of this document, BHA emergency awards may include award-specific monitoring, evaluation, and reporting requirements. Partners should thoroughly review their award documents and coordinate with the AOR to ensure that they fulfill all requirements.

1.3 APPLICATION REQUIREMENTS

1.3.1 M&E PLAN

The purpose of the M&E Plan is to serve as a framework for activities to demonstrate accountability and improve the quality of activity implementation and outcomes for participants. The M&E Plan should serve as a roadmap for activity staff, documenting M&E approaches and processes in sufficient detail. It should demonstrate that a partner has a rigorous system for monitoring and evaluating activity performance in a way that produces accurate, reliable, and useful data in a timely manner for decision-making.

BHA requires the submission of an M&E Plan as an annex to the application. The following two components should be included in the M&E Plan:

- Component 1: Indicator Tracking Table (ITT)
- Component 2: M&E Plan Narrative
    - Monitoring Approach including Abbreviated SOW for Baseline/Endline study, if applicable
    - Evaluation Approach including Abbreviated SOW for Evaluation, if applicable

The components of the M&E Plan due at application should be submitted as two attachments: a Word document (M&E Narrative) and an Excel document (ITT). Suggested formats for the ITT are included in the BHA Emergency M&E Templates. Applicants are encouraged to use these suggested formats when developing their M&E Plans, but may use other formats as long as the required information is included. M&E Plans must be submitted with full applications, but are not required with concept notes. A suggested outline for the M&E Plan narrative is included in Annex 1.

The M&E Plan should also include a description of M&E staffing and resources, including a summary of the M&E budget. BHA encourages partners to budget at least three percent of the total budget to M&E. This may vary slightly by award size, with larger-budget activities spending a smaller percentage, and smaller-budget activities spending a higher percentage. BHA encourages you to include an M&E Specialist or equivalent position, as well as costs associated with data collection and resources, in the staffing plan and budget. Include an explanation of the M&E staffing plan and associated costs, including for Information and Communication Technology (ICT).

The technical guidance in this document is applicable to all BHA non-PIO emergency awards issued under the BHA Emergency Application Guidelines. Figure 1 summarizes the key components of the M&E Plan throughout the award cycle and the associated submission timing requirements.

The ITT serves as a means to articulate and monitor the progress of the intended results of the activity and illustrate its integrated logic (See Chapter 2 for additional Indicator Tracking Table information). Indicators that are included in the ITT will be used to track progress of the activity and are required regardless of duration. The logical structure of the ITT is designed to organize an activity by purpose and sub-purpose(s), and should, at a minimum, consist of the following components:

- Results Statements;
- Indicators;

- Data Sources/Methods;
- Targets; and
- Assumptions

Targets must be provided for all indicators for the life of award (LOA). Applicants are required to include all required (R) and required-if-applicable (RiA) or other standard BHA emergency indicators. For more information about BHA emergency indicators, please refer to the BHA Indicator Handbook for Emergency Activities. Because the baseline values are not available at the application stage, the final targets may be expressed in relation to the baseline value (e.g., "baseline + 10 percentage points"). After a baseline study is completed, partners should document any updates made to targets in (1) the Baseline Report submission, which includes an updated ITT, and (2) in the Award Results Tracking System (ART). See Chapter 2 for additional guidance on the ITT.

The M&E Plan Narrative allows applicants to outline their approach to monitoring and evaluation specific to the context of the activity. Moreover, applicants are encouraged to provide a detailed plan for their staffing and allocation of resources for the monitoring and evaluation component of their activity. The M&E Plan Narrative includes the Monitoring Approach and Evaluation Approach.

The Monitoring Approach includes a description of the type of monitoring, indicators, methods, and the data collection, quality, management and safeguarding procedures and resources that the partner will use during the course of planning, implementation, and evaluation. See Chapter 4 for further guidance.

The Evaluation Approach should include a narrative that describes the evaluation If no evaluation is planned, the Evaluation Approach should note that, provide rationale, and describe what assessments of any kind are planned. See Chapter 6 for further guidance regarding the evaluation structure and content.

PIRS for all custom indicators are to be submitted within 90 days of the award, and may be annexed to the baseline report submission. Moreover, BHA recommends, but does not require, submission of PIRS for all contextualized standard BHA indicators. Please see Chapter 2 for further guidance.

1.4 POST AWARD M&E DELIVERABLES

The initial three months of the award are a critical period for monitoring and evaluation. During this time, partners should refine and finalize indicators and targets, conduct the baseline study (required for awards of six months or more), develop and refine their monitoring system and tools, and plan procurement for an evaluation (required for awards of 18 months or longer).

1.4.1 BASELINE/ENDLINE REPORT

A baseline data collection is required for all non-PIO BHA emergency awards that are six months or longer in duration. The purpose of the baseline study is to collect baseline values for specific outcome indicators that will be compared to values collected at the endline and to provide information to the partner about the activity's target population to strengthen the design and targeting of interventions. Note that in general, BHA does not encourage partners to conduct representative surveys at

baseline/endline for activities less than 12 months in duration. Partners that conduct baselines must submit the baseline report, an updated ITT, final targets, and PIRS for all custom indicators to the AOR and uploaded into BHA's Awards Results Tracking System (ART) within 90 days of award approval. Related data sets must be submitted to the DDL, in accordance with ADS 579, before the closeout of the award. Partners must also submit the updated ITT, with the baseline values and final targets, as part of the AR at the end of the fiscal year in which the baseline survey was completed. Chapter 5 provides detailed guidance about the baseline study.

1.4.2 SEMI-ANNUAL REPORTING

Partners must fulfil their annual reporting requirements by submitting a Semi-Annual Report (SAR) no later than April 30. Please refer to section 7.1.1 for more information.

1.4.3 ANNUAL REPORTING

Partners must fulfil their annual reporting requirements by submitting an Annual Report (AR) no later than October 30. Please refer to section 7.1.2 for more information.

1.4.4 FINAL PROGRAMMATIC REPORTING

Final performance reports (FPR) are due 90 days from the award end date. Final performance information must be reported at the end of the activity life for the entire life of the activity. Please refer to section 7.1.3 for more information.

1.4.5 EVALUATION SOW AND REPORT

Evaluations are required for emergency activities that meet one of the two criteria outlined in the BHA Emergency Guidelines: 1) if the original period of performance for the activity is 18 months or longer, or 2) if your organization has implemented at least one BHA-funded award (of any duration, in any sector) in the past three years in a given country and your organization has not completed an evaluation of any BHA-funded awards in that given country in the past three years[1]. Final evaluations must be conducted by an internal team led by an experienced team leader, who is external to the organization, or by an external firm. Partner staff who are not substantially engaged in the design or implementation of the activity under evaluation may participate in the evaluation. USAID staff may also participate in the evaluation. Activities with smaller budgets may opt to hire an individual consultant to oversee the final evaluation, with baseline and endline data collection conducted by project staff, while activities with larger budgets may opt to hire an external firm to conduct the entire final evaluation including endline data collection.

While not required by BHA, awards less than 18 months can plan for a final evaluation to capture best practices and lessons learned. **If a partner plans to conduct an evaluation, an abbreviated SOW must be submitted with the application regardless of the duration of the activity**. The abbreviated SOW must include the following sections: *Evaluation Objectives*, *Illustrative Evaluation Questions/Topics*, and *Evaluation Methodology*. Additional guidance for the SOW can be found in Annex 3.

[1] This second criterion for evaluation applies to BHA activities awarded on/after October 1, 2020; the three year timeframe is not inclusive of former-OFDA and former-FFP awards.

In some instances, BHA may choose to contract and manage an evaluation directly. In such an event, the AOR will notify the partner at least six months prior to the end of the activity.

For partner-managed evaluations, the partner must submit the final report and related documents to the DEC and related data sets to the DDL within 90 day before the award end date. (See Chapter 7 for further guidance and information regarding reporting.)

1.4.6 M&E PLAN FOR COOPERATIVE AGREEMENTS

The above post-award deliverables are required for all BHA assistance mechanisms - including grants and cooperative agreements. Partners awarded a cooperative agreement that includes "substantial M&E involvement" as part of the award provisions must also submit an updated M&E Plan within 30-60 days of the start of the award to the AOR. BHA recommends partners review their terms of their award closely to confirm the timing of submission for cooperative agreement deliverables.

The M&E plan submitted as a cooperative agreement deliverable may be based off of the M&E Plan submitted at application, but should include additional detail and/or any new information from the partner. Annex I provides a suggested format for the M&E Plan, which can be referenced in the development of the Application and post-award M&E Plan Deliverable submissions.

1.5 MONITORING & EVALUATION SECTOR

The M&E Sector is distinct from, and does not replace, the M&E Plan annex which is required for all BHA applications. The M&E Sector is designed to capture operational research related to M&E, and consists of two sub-sectors, (1) Advancing Evaluation for Humanitarian Assistance, (2) Monitoring & Data Utilization. Indicator requirements related to this sector would be most appropriate for activities focused on M&E operational research.

As a result of the knowledge that is generated through the M&E Sector, BHA has the capability to systematically keep track of best practices related to M&E, and can serve to strengthen other partners' M&E systems. The overall objective is to support the humanitarian community's commitment to invest in initiatives that will improve M&E practices.

CHAPTER 2: INDICATORS AND INDICATOR TRACKING TABLE

2.1 ACTIVITY OVERVIEW AND DESIGN

As part of BHA's Grand Bargain Commitments and in response to partner feedback, BHA has integrated elements of a Logical Framework into the Indicator Tracking Table (ITT). The ITT is useful for both managers and M&E staff throughout the program cycle to articulate the intended results of the activity and how it will be monitored. The ITT incorporates the results hierarchy (Goal, Objective, Intermediate Outcome, Outputs) of a logic model that provides a description of how an activity is to function in the form of a linear chain of cause and effect. When designing the ITT, it is important to consider the theory of change underlying the activity design. This can significantly improve the logical coherence and the soundness of activity design, and help to identify the assumptions that are critical to the success of an activity.

The first step to designing an activity and identifying intended results is to conduct a problem analysis. Applicants need to not only understand the immediate needs of the affected population, but also identify what the root causes of those issues are in order to design the most effective response. For example, the proper response to food insecurity driven by drought may be very different than to food insecurity driven by a conflict that disrupts markets. Applicants should use both primary information (i.e., needs assessments) and secondary information (Famine Early Warning Systems Network (FEWS NET), OCHA Humanitarian Response Plan (HRP), Standardized Monitoring and Assessment of Relief and Transitions (SMART) Survey reports, etc.) to identify problems that lead to humanitarian crisis.

Once the applicant has a well-defined problem, they can begin developing the hypothesis - or a Theory of Change - to understand the set of interventions required to change the conditions, practices, or behaviors, and eventually address the main problem. The proposed interventions may not address all of the conditions required to achieve the overarching goal but must demonstrate contribution. The Theory of Change does not require an extensively detailed narrative or supporting diagrams but must be developed using sound evidence. Should there be any gaps in evidence, the applicant should plan to use rapid data collection tools to fill the evidence gaps. Please refer to the BHA Emergency Application Guidelines Activity and Design section for more detailed theory of change requirements.

Finally, the assumptions underlying the theory of change should be identified and assessed to determine the feasibility of the selected approach. For example, if a critical assumption is unlikely to hold, then the approach should be reconsidered. These assumptions will be documented in the ITT, and should inform the activity's monitoring strategy. Assumptions beyond the control of the applicant and necessary for the achievement of objectives at all result levels (e.g. the exchange rate remaining consistent) should be monitored throughout the life of the award (*see Section 4.2: Context Monitoring*).

2.2 INDICATOR TRACKING TABLE FORMAT

Box 2. Summary Indicator Tracking Table Requirements

When to submit the Indicator Tracking Table

- At Application: Required for all applications regardless of duration
- Post-Award: Submit an updated ITT within 90 days of award approval that includes actual baseline/base values for all indicators and any updates to indicator targets.
    - If the partner is submitting a Baseline Report, the updated ITT should be submitted as an annex to the Baseline Report

When to submit PIRS

- Post-Award: PIRS for all custom indicators in the Indicator Tracking Table must be submitted within 90 days of award approval.
    - If the partner is submitting a Baseline Report, PIRS should be submitted as an annex to the Baseline Report
- BHA recommends but does not require that you submit PIRS for all contextualized standard BHA indicators.

Required indicators

- Ensure that all required, required select 2 or 3, and required-if-applicable or selected optional BHA indicators indicators are included in the Indicator Tracking Table (*see Section 2.3*)

The Indicator Tracking Table documents the results statements in the proposed results hierarchy, associated sector(s), sub-sector(s) and keyword(s), the indicators, the disaggregates, indicator type, desired direction of change, targets, actual values, data methods/sources, data collection frequency, position responsible for each indicator and assumptions. The ITT is organized by information required during application submission, as illustrated in Figure 1 and data collected during implementation as illustrated in Figure 2. The suggested format for the BHA Indicator Tracking Table is included in the BHA Emergency Resource Page.

Figure 1: ITT Template

Result Statement	Sector	Subsector	Keyword	BHA Indicator No. ^	Indicator^	Disaggregates^	Indicator Type ^	Desired direction of change (+/-) ^	Data Collection Method^^	Data Source^^	Data Collection Frequency ^^	Position Responsible for Collection	Assumptions	Comments
Goal														
Purpose														
Sub-Purpose*														
Intermediate Outcome*														
Output														
Input^^^														

Figure 2: ITT Template- Monitoring Data

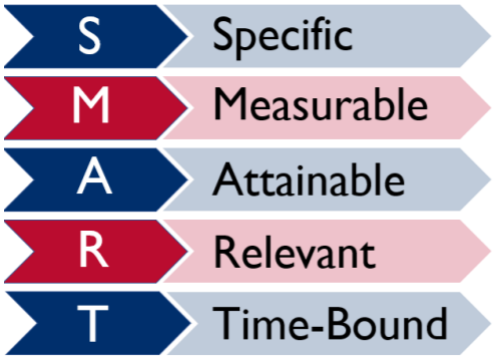
Sector	Indicator^	Baseline Value	Endline Value	Monitoring Data**						
				Life of Award (LOA) Target	Actual (Semi-Annual 1)	Actual (Semi-Annual 2)	Actual (Year 1)	LOA Actual	Variance	Comments

Every application must include an Indicator Tracking Table. The ITT details key elements of an activity under several columns and rows.

Goal: The highest-level, long-term result to which an activity seeks to contribute. The Goal aligns with BHA's mission and the goal of the humanitarian response. Typically, a Goal cannot be fully accomplished by the activity during the award period. Factors beyond the control of the activity must also be addressed before the goal can be fully accomplished.

Purpose: A key, high-level result that the activity is accountable to accomplish during the LOA. A Purpose describes a desired change in the condition or status of the population in the target area to which the outputs and outcomes of the activity's interventions should contribute significantly.

Sub-Purpose: A result of one component of the activity necessary for a Purpose to be achieved. The Sub-Purpose statement must be SMART. These often include behavioral and systemic changes, for example, adoption of promoted techniques or behaviors. Including Sub-Purposes is **optional** for a single-sector activity or activities that have multiple, non-integrated sector purposes. Partners must include Sub-Purposes for complex integrated and multi-sector purposes.



Intermediate Outcome: An outcome that must occur before a Sub-Purpose or a Purpose can be achieved, such as changes in knowledge or attitudes, mastery of skills, and adoption of new methods. There may be multiple levels of Intermediate Outcomes in sequence along a single pathway. Including Intermediate Outcomes is **optional**. Partners may choose to include Intermediate Outcomes depending on the complexity of their activity design.

Output: An output is a tangible, immediate product of an intervention under the activity's control or influence. Examples include "Food vouchers provided to target households," "infant and young child feeding (IYCF) training provided to mothers groups," or "ready-to-eat rations distributed to displaced households."

*Note: **Input** indicators are not required in the Indicator Tracking Table with the exception of BHA Indicators [H26, P2, P4 and S13 as applicable].*

The Goal, Purpose, and Output levels of the ITT are **required** for all BHA emergency applications. The Sub-purpose and Intermediate Outcome layers are **optional** depending on the complexity of the activity being proposed. Applicants should decide whether or not the additional layers of the ITT are necessary to effectively communicate the activity's design and monitor implementation.

A suggested format for the ITT and an example is included in the BHA Emergency M&E Resources Page. Detailed definitions of BHA Indicator Tracking Table Columns can be found on the "Definitions" tab of the suggested format. It is meant to be a starting point for partners and should be adapted to match the activity's design.

2.3 INDICATORS

BHA tracks two primary types of indicators: 1) performance indicators, and 2) context indicators. Performance indicators are used to measure whether or not the outputs and outcomes in the ITT are being achieved. Context indicators (discussed in detail on page 12) are used to measure external factors that are relevant to the success of the activity (i.e. the assumptions in the ITT). At least one performance indicator should be included for each output and outcome. Targets should be included for each performance indicator, indicating what will be achieved over the life of award (LOA).

The ITT should include the following indicators:

- All R, R-Select 2 or 3, RiA and selected optional BHA indicators
- Custom indicators selected by the applicant
- Context indicators may be optionally included in the ITT. If included, they should be placed in the *Assumptions* column.

2.3.1 INDICATOR TYPES

BHA Indicators

Please refer to the *BHA Indicator Handbook* on the BHA Resources page. The BHA Indicator Handbook includes the PIRS for all of the BHA standard indicators, with details on the indicator definition, data collection, and indicator calculation.

Custom Indicators

Applicants are encouraged to create custom indicators to measure specific activity outputs, outcomes, and context for which there are no corresponding BHA indicators, with preference to use of indicators from the IASC Emergency Indicator Registry. Custom indicators may also be adopted from the Office of U.S. Foreign Assistance Resources (F) Standard Foreign Assistance Indicators, from the Infant and Young Child Feeding in Emergencies Operational Guidance, from other external groups (e.g., United Nations (UN) Specialized Agencies, other donors, or the Sphere Handbook), or they may be created by the activity's M&E personnel. Any indicators that are internal to their organizations (e.g., key performance indicators used on all donor-funded awards or contracts) should be labeled as custom indicators in the

M&E Plan. The PIRS must be submitted for all custom indicators. All complementary nutrition activities must have at least one outcome indicator. Partners are encouraged to consider the total number of indicators and the costs associated with their measurement when deciding to add new custom indicators.

Context Indicators

There are factors outside of the control of every activity that can affect whether or not the outcomes are achieved. These context indicators can be tracked in the ITT. For example, an activity that provides cash to enable target households or individuals to pay for rent or purchase shelter necessities may require housing availability and price stability and product availability in local markets in order to achieve safe shelter outcomes. Context indicators are often identified as risky assumptions or assumptions which have the highest level of uncertainty. BHA recommends that applicants define custom context indicators that are important to monitor the activity and understand the intervention's results. Actual values for context indicators can be reported in the Indicator Tracking Table, but no targets are required. BHA expects partners to define their own custom context indicators that are relevant to their specific operational environment. Context indicators may vary substantially between partners.

2.3.2 INDICATOR TARGETS

A target is a measurable value that represents a specific, planned level of achievement to be accomplished (output) or a change that should occur (outcome) within a specific timeframe. Typically, indicator targets for emergency activities will be for the life of award (LOA). Targets should be included for both output and outcome indicators. No targets are required for context indicators, but they can be useful to set thresholds upon which an action will take place (e.g., re-evaluate voucher value once inflation reaches a certain level; triggering changes in security protocols if conflict increases).

Targets serve multiple purposes:

1. **Establish shared goals**
    - Give stakeholders a common understanding of what to expect from the activity
    - Provide justification for the investment
    - Help to measure effectiveness of the proposed interventions
2. **Monitor progress**
    - Provide benchmarks for accountability
    - Provide evidence whether the theory of change is working
    - Promote transparency
3. **Learning**
    - Give insights into what should be adjusted in future activities

Targets should be **ambitious** yet **achievable**. They should motivate partners to "reach" while also being realistic. The basis of the targets should be rational. Targets must be consistent with the underlying logic of the activity design, and with time and budget constraints.

Population vs. Beneficiary Targets

When setting targets, it is important to determine whether the measurement will take place at the population/community level, or the participant level. Most BHA emergency activities will measure indicators at the beneficiary level, either through a beneficiary-based survey or collection from all participants (e.g., census). Participant-based measurement of indicators makes target-setting simpler.

A population-based survey may be required in cases where the interventions are at the community-level (e.g. water point rehabilitation), or if the intervention is designed to have a population-level effect through secondary adoption (which is rare for emergency activities). For example, the indicator "Percent of households in target area practicing open defecation" would require a population level measurement that would include members of the target area which may not have been reached by the intervention. Population-based measurements are also used when assistance is available and accessible to the entire population in the intervention area. When setting population targets, it is important to consider the baseline value, the coverage of the intervention, the timing and duration of the activity, and the effectiveness of the intervention.

For example, when setting a population-level target for the indicator "Percentage of households with poor, borderline, and acceptable Food Consumption Score (FCS)" consider:

- Baseline value: what percent of households currently fall into each category in the target areas?
- Saturation: what percent of households in the target areas will be reached by the intervention?
- Effectiveness: what percent of households reached are expected to be in the 'acceptable' food consumption category after the intervention?

2.3.4 PERFORMANCE INDICATOR REFERENCE SHEET (PIRS)

A PIRS is a tool used to define indicators. PIRS are important for ensuring indicator data quality and consistent approaches to data collection. A well-designed PIRS should be clear enough that if the M&E Manager is to abruptly leave, the successor could continue measuring and reporting the activity's indicators in a consistent manner without ambiguity. Since both BHA and the applicant's headquarters aggregate data collected by different activities in different countries for reporting and analyses, PIRSs help to ensure the consistency of data for a specific indicator. Variation in indicator definition, disaggregation, or computation will limit the ability to aggregate the data.

The objective of a PIRS is to describe the indicator in detail, which should include:

- What raw data are needed
- What survey questions to ask, or observation processes to follow to get accurate raw data
- Who is responsible for collecting the data
- Which tools will be used for data collection
- From whom will data be gathered, or what will be observed
- Precisely when data will be collected
- How the collected data will be used to calculate the indicator value
- In what unit the indicator will be presented in
- What disaggregations will be reported
- Definition of all terms in the indicator

PIRS for all BHA emergency indicators, including the PIRS template are in the BHA Indicator Handbook. These PIRS should be used to ensure that the indicators are measured consistently across partners. If necessary, the BHA emergency PIRS can be contextualized to meet the specific needs of the partner and the context in which they are operating. These changes should not alter the underlying definition or calculation of the indicator, and all changes should be clearly documented.

PIRS for all custom indicators in the activity's ITT are due within 90 days of award approval along with the baseline report. BHA recommends but does not require that the partners submit PIRS for all contextualized standard BHA indicators. For custom indicators, partners are required to develop their own PIRS following the BHA template so that BHA can understand what the indicator is measuring and how it will be calculated.

CHAPTER 3: DATA COLLECTION METHODS

The following sections provide guidance on data collection methods, standards, and practices to be used in the monitoring and evaluation of emergency activities. All data collected under the M&E system should be actionable, and the methods used to collect the data should be determined based on when the data is needed to support decision-making. Given the short life of most emergency activities, and complex operating environments, it is critical that activities only collect data that will be useful. The M&E Plan should not only document how data will be collected, but also specify how it will be used. There are two components to creating actionable data. First, the organizational use of each piece of data should be articulated, including reporting, learning, and management decision-making. Second, monitoring systems should be designed to ensure that information gets to those who need it.[2]

Table 2. Data Collection Methods by Purpose

<table>
<tr><th rowspan="3"></th><th colspan="6">Purpose of Data Collection</th></tr>
<tr><th rowspan="2">Baseline Assessment or Endline</th><th colspan="3">Performance Monitoring</th><th rowspan="2">Context Monitoring</th><th rowspan="2">Evaluation</th></tr>
<tr><th>Output</th><th>Outcome</th><th>Process</th></tr>
<tr><td>Beneficiary Based Survey**</td><td rowspan="2">As applicable*</td><td>Not applicable</td><td rowspan="2">Recommended: Sampling design using SRS, 2 stage cluster, or PPS</td><td>SRS, 2 stage cluster, or PPS</td><td>Not recommended</td><td rowspan="2">Recommended or Required***</td></tr>
<tr><td>Population Based Survey**</td><td>Not applicable</td><td>Not recommended</td><td>Not recommended</td></tr>
<tr><td>Census</td><td>Not recommended</td><td>As applicable</td><td>Not recommended</td><td>Not recommended</td><td>Not recommended</td><td>Not recommended</td></tr>
<tr><td>Routine Monitoring</td><td>Not applicable</td><td>Recommended</td><td>Recommended</td><td>As applicable</td><td>Recommended</td><td>As applicable</td></tr>
<tr><td>Secondary Data Collection</td><td>As applicable</td><td>As applicable</td><td>As applicable</td><td>As applicable</td><td>As applicable</td><td>As applicable</td></tr>
<tr><td>Qualitative**** Methods</td><td>Recommended</td><td>Not Recommended</td><td>Recommended</td><td>Recommended</td><td>Optional</td><td>Recommended</td></tr>
</table>

*For awards less than 12 months, BHA does not recommend conducting representative surveys.
**Pre-post minimum sample size estimation with SRS, 2 stage cluster, or PPS method is recommended.
***For awards 18 months, an evaluation is required. Pre-post minimum sample size estimation with a representative sample using SRS, 2 stage cluster, or PPS method is required.
****Qualitative methods may include key informant interviews, focus group discussions, and/or other methods.

[2] Adapted from "Monitoring for Learning and Accountability", Goldilocks Toolkit, Innovations for Poverty Action (2016).

3.1 METHODS

There are a variety of data collection methods that may be used to generate information about an activity's performance and/or the operating context. Table 2 below summarizes which methods are recommended for baseline/endline studies; the three types of performance monitoring discussed above; context monitoring; and evaluations.

3.1.1 BENEFICIARY BASED SURVEY (BBS)

Beneficiary based surveys are conducted among a sample of beneficiaries that participates in an activity's interventions. In the context of emergency activities, BBS is commonly used to collect baseline data, post-distribution monitoring, and endline data. Beneficiary based surveys typically use questionnaires to gather information from a probability sample of individual beneficiaries or beneficiary households. A probability sample ensures that every individual or household from the entire pool of beneficiaries has an equal likelihood of being selected in the sample. The sampling frame only includes beneficiaries and the sampling design should ensure that a minimum number of individuals or households are included in the survey to ensure results of the survey are representative of the entire cohort of beneficiaries with the desired level of precision. When possible and appropriate, BHA typically recommends beneficiary-based surveys rather than population-based surveys. Finally, some partners collect baseline data during the registration process using a systematic sampling method, such as conducting the full baseline survey with every Nth person to complete the registration process.

Note that some beneficiary-based surveys may in fact constitute the entire cohort (or population) of people or households in a given area, e.g., an IDP camp. This is referred to in the BHA Indicator Handbook as a "beneficiary based survey" or "beneficiary/population based survey" since all members of the population/cohort are beneficiaries.

Note that direct observation may be a useful method for verification of data during surveys, e.g., verification of latrines.

3.1.2 POPULATION-BASED SURVEY (PBS)

Population-Based Surveys use questionnaires to gather information from a probability sample of all individuals or households in a given area, typically the entire area of implementation for an activity. A probability sample ensures that every individual or household from the entire survey population (i.e., all people or all households in the area of implementation) regardless of their participation in activity interventions has an equal likelihood of being selected. The sampling frame includes all individuals or households in the area and the sampling design should ensure that a minimum number of required individuals or households are included in the survey to ensure results of the survey are representative of the entire cohort of beneficiaries with the desired level of precision. BHA only recommends using population-based surveys when the interventions can potentially benefit the entire population and indicator estimates cannot be generated based on beneficiary-based survey data. PBSs may be necessary when interventions benefit the entire community and do not have a defined beneficiary list from which to sample (e.g. borehole rehabilitation, hygiene promotion, etc.).

3.1.3 CENSUS

A census typically involves using a checklist or questionnaire to gather information from or about all entities (e.g., people, households, water points) within a given activity or intervention. BHA generally does not recommend using a census to gather information for outcome indicators from large cohorts of beneficiaries or beneficiary households. However, some partners choose to collect baseline data as part of the registration process, which is considered a census (since data are collected from every individual or household). When collecting baseline data at registration, BHA encourages partners to only collect baseline data from a sample of beneficiaries. Note that a limited number BHA PIRS require a census, typically of hardware such as water points.

3.1.4 ROUTINE MONITORING METHODS

Routine monitoring refers to data that is collected on an ongoing basis by activity staff throughout implementation. Routine monitoring data is typically collected from **some or all direct beneficiaries**, and measures indicators at the output and outcome levels. Routine monitoring requires staff to allocate time so that they can collect data from beneficiaries, and for M&E staff and supervisors to regularly review and spot-check that data to identify issues. This system of checks instills confidence in the integrity of the data, thereby allowing the activity to use the data in near real time to review progress and identify challenges. Routine monitoring methods may include, but are not limited to, the following:

- Direct observation (e.g., staff use checklists to systematically record observations about practices or conditions on the ground during a field visit; staff keep records from a food or voucher distribution verifying the participant ID and ration received; or staff keep transaction records of a transfer program using ATM cards or mobile-money transfers).
- Compiling sign-in sheets or other trackers from training.
- Document review/audit (e.g., reviewing water user committee documents).
- Pre and post knowledge tests from training activities to measure knowledge retention.
- Diaries, whereby activity beneficiaries, community mobilizers, or frontline staff/volunteers are trained and given a notebook to record practices in writing or pictures; these data are typically verified by activity staff then copied to the activity database.
- In rare circumstances it may be appropriate to hold regular focus groups to monitor conditions on the ground. For example, the indicator "Percent of water user committees created and/or trained by the WASH activity that are active at least three (3) months after training" would only be possible if focus group discussions from the active water user committees were held.
- A survey to test for water quality at water access points constructed under the activity with BHA funds.

3.1.5 SECONDARY DATA

Emergency activity partners may use both primary or secondary data in their monitoring and evaluation. Primary data refers to data that is collected directly by the partner, e.g. using routine monitoring or survey methods. Secondary data are data collected by someone else for a different purpose. The data could have been collected from other sources, such as host country governments, the cluster system, or other partners. While primary data is preferred, the use of secondary data should be explored, especially for context monitoring, including market monitoring.

3.1.6 QUALITATIVE DATA METHODS

Qualitative data collection methods such as key informant interviews or focus groups may be used for process monitoring such as quality of behavior change sessions or demonstration plots, outcome monitoring such as women's empowerment, context monitoring such as conflict dynamics, unintended consequences, magnitude of inclusion and exclusion errors, and secondary adoption of promoted behaviors/practices. Qualitative assessments may be used to answer discrete questions that arise during implementation, provide explanations for patterns in quantitative data, or inform specific strategies.

Qualitative methods may be useful for monitoring and/or evaluating the following:

- **Outcome monitoring:** There are anticipated outcomes that are not easy to quantify, therefore, qualitative tools and methods are suitable to capture these outcomes. For example, peoples' agency, empowerment, gender equity in decision making, coping strategies, and changes in norms and attitudes.
- **Process monitoring:** Monitoring of implementation processes such as training, behavior change sessions, distribution of food and non food items, and construction work, can help identify sub-optimal quality of implementation which will hinder activity performance. Direct observation of training sessions, discussions with the beneficiaries, interviews with front line staff, site visits to beneficiary homes, clinics, and other locations may be used.
- **Post distribution monitoring:** Qualitative methods are useful to understand protection issues, transaction costs, and waiting time, among other factors. It may also be more appropriate in specific settings (e.g., school feeding programs or sensitive contexts where surveys are not feasible).
- **Unexpected and unexplained achievements:** Quantitative indicators may suggest that progress toward a quantitative target is not on track (e.g., when progress against targets is unexpectedly low or high). Qualitative methods or tools could be used to understand the reasons behind this under- or over-performance. The information then can be used to tailor the implementation strategy either to improve performance or use it as a positive deviance to inform other interventions.
- **Unintended effects:** Qualitative data collection is well-suited to explore possible unintended consequences or unexpected outcomes of interventions that would be overlooked in routine quantitative monitoring.
- **Secondary Adoption:** In some instances, BHA emergency activities may be designed to affect change at the population level by directly engaging with a cohort of households or communities who will subsequently share the key knowledge/skills/practices/resources at the population level. Qualitative methods may be appropriate to monitor secondary adoption (by non-beneficiary community members) to see if there are observable, population-level changes. Qualitative methods may help to get a sense of the magnitude of secondary adoption and understand why certain practices are adopted by neighbors and what could be done to further promote secondary adoption.

3.2 SAMPLING GUIDANCE FOR PROBABILITY-BASED SURVEYS

This section provides practical guidance on the steps for sampling in probability-based surveys:

1) how to identify and construct a sampling frame,
2) how to determine what kind of sample design to use, and
3) how to determine the appropriate sample size. The last section provides guidance on sample weighting and data analysis techniques.

This guidance applies for activities that report on indicators measured through probability-based surveys, and provides best practice to ensure surveys are appropriately designed to be statistically valid and representative of the target population or beneficiary cohort. BHA does not expect that every activity will conduct statistically comparable baseline/endline surveys.

BHA requires probability-based baseline/endline surveys for a limited set of indicators. In some cases, monitoring indicators measured through post-distribution monitoring (PDM) surveys may also employ probabilistic sampling. **Partners should review the PIRS for their required indicators to determine whether a representative baseline/endline or monitoring survey is required**. If an indicator requires representative surveys, but such data collection is not feasible due to prohibitive operational constraints (e.g. data collection may endanger beneficiaries or staff), the partner should propose strong justification to omit this indicator in the Application M&E Plan, for BHA review and approval.

3.2.1 SAMPLING FRAMES

Sampling is an efficient way to identify a subset of the survey population which can be used to provide estimates of characteristics and/or indicators for the entire target population. A well-designed sample saves time and resources while still generating precise information about the full target population.

A **sampling frame** is a group of units from which a subset (sample) is drawn (e.g., all beneficiaries of an activity or all beneficiaries receiving conditional transfers or all health clinics covered by an intervention or all health clinics in a country).

Regardless of whether the purpose of the survey is for baseline/endline data collection or for monitoring, the first step in designing a probability-based survey is to define the survey population of interest. The survey population can be referred to as the sampling frame. These are the groups or groups of people/households from which the sample for the survey will be drawn. There may be only one group of individuals targeted by an activity, or multiple groups of people such as farmers, children, and women at reproductive age targeted by an activity. These target groups may be distinctly separate from each other such as children and women at reproductive age or may overlap such as farmers and women in reproductive age. Survey populations are defined based on the population targeted by various interventions and the set of indicators the activity is planning to measure. Table 3 below provides examples of indicators and the target population for each.

Table 3. Select BHA Indicators and Respective Sampling Frames

Indicator	Sampling Frame
A02. Number of hectares under improved management practices or technologies with BHA assistance	Individual beneficiary farmers who received training
A12. Percent of individuals who received training that are practicing appropriate crop protection procedures	Individual beneficiary farmers who received crop protection training
F02. Percent of households where women reported participating in decisions on the use of food assistance	Beneficiary households with women
M03. Percent of beneficiaries reporting that humanitarian assistance is delivered in a safe, accessible, accountable, and participatory manner	Individual beneficiaries
N8. Percent of infants 0–5 months of age who are fed exclusively with breast milk	Infants 0-5 months in beneficiary households
N09. Percent of children 6–23 months of age who receive foods from 5 or more food groups	Children 6-23 months in beneficiary households
N10. Percent of women of reproductive age consuming a diet of minimum diversity (MDD-W)	Beneficiary women of reproductive age
W4. Percent of households targeted by the WASH promotion activity that are properly disposing of solid waste	Households targeted by the WASH promotion activity
W15. Percent of households in target areas practicing open defecation	All households in implementation area

3.2.2 SOURCES FOR SAMPLING FRAMES

Once the survey populations are defined, you can construct sampling frame(s). In most cases, sampling frames for beneficiary-based surveys can be constructed from the beneficiary registry of households or individuals. Most emergency activities record/register households or individual beneficiaries (depending on the targeting and intervention strategy). For example, if different interventions are targeted to different beneficiaries using a beneficiary registration system, a beneficiary register is the best source of information to construct a sampling frame or frames because it should perfectly reflect the survey population. By contrast, for a community-level intervention that targets all community members (i.e., when it does not make sense to generate and maintain a beneficiary registry), a population-based survey would be more appropriate.

Partners should design a population-based survey when the interventions are designed to benefit entire communities. For population-based surveys, all households or individuals in the target communities or implementation area are considered as survey populations. To minimize the cost and logistical burden of population-based surveys, sampling frames are typically constructed at two levels: the first is the community/village level and the second is the household level. A list of communities/villages in the target

area can often be provided by the activity or identified through census files available from the last official census taken in the area. Once a sample of communities is drawn, then a list of households in the selected communities may be available through local community authorities or other community groups. If household lists are not available, households can be sampled using systematic sampling on the ground (see Section 3.2.3).

3.2.3 DATA TO BE INCLUDED ON SAMPLING FRAMES

Beneficiary household/individual level sampling frames should include the following key elements:

- Unique household identification number or unique beneficiary identification number (depending on the targeting strategy)
- Contact information (including name, physical location, primary phone number [if available], and secondary phone number [if available]).
- When possible, household characteristics (household gender composition, size, primary and secondary livelihood activities)
- Intervention(s) received

If all the relevant information listed above is recorded in the database, during beneficiary registration, this information does not need to be collected again in the endline survey. Ultimately, an investment in data collection at the time of registration will increase the efficiency and improve the quality of the survey data and analysis by limiting interviewer and respondent burden and providing additional covariates for use during analysis. Community/village level sampling frames should include the name of the village, GPS coordinates if available and higher-level geographic identifiers such as department, region, commune, etc. A measure of size, either population or number of households in each village is needed in order to use probability-proportional to size sampling (see Section 3.3.3). This type of information can be obtained from community level records or census data.

3.2.4 TARGET GROUP BY ACTIVITY INTERVENTION

Based on the targeting strategy, a baseline survey design that requires multiple sampling frames must organize the target groups that will receive a similar set of interventions. Partners may want to use Table 4 below to assist with identifying sampling frames and sample sizes using the estimated numbers that were used to develop the interventions and budget. If multiple sampling frames are needed, applicants should identify the key indicators for each sampling frame and calculate sample size for each indicator. For each sampling frame, a separated survey should be conducted. Conducting surveys with multiple sampling frames is complex, so partners should ensure that they have sufficient technical support.

Table 4. Sampling Frames and Sample Sizes

Intervention [e.g. Cash transfer, seed inputs]	Indicator [e.g. Percent of households with access to sufficient seed to plant]	Target Group [e.g. all households, women of reproductive age]	Target Beneficiary Number

Questions to Ask when Developing Sampling Frames:[3]

1. ***Which group of people are expected to receive benefits from this intervention?*** Knowing the target groups for your study indicators will help determine the appropriate sampling frames.
2. ***What is the sample frame for each stage of sampling?*** For one-stage beneficiary-based surveys, sampling frames are typically beneficiary registries. A multi-stage cluster design may be more appropriate for a population-based survey. For multi-stage designs, construct a sampling frame for each stage.
3. ***How is the sampling frame being constructed?*** Identifying good sources for sampling frames can be difficult and many sampling frames are only proxies for the entire survey population. Obtaining a full list of beneficiaries is important for establishing a representative sampling frame.
4. ***What are its limitations in generalizing to the study population?*** If a sampling frame does not include everyone who is supposed to be benefiting from an intervention or a set of interventions, the survey results will not be representative to your population of interest. This can happen because of safety and security concerns or limited access due to seasonality or other factors. This is a limitation that must be noted in the study report and considered when interpreting the results.

3.2.5 SAMPLING FRAMES – PRACTICAL EXAMPLES

In developing sampling frames, an activity needs to take into account its targeting strategy and the (possibly different) target populations for various interventions. Box 3 provides guidance for three scenarios with different targeting strategies and target populations.

Box 3. Sampling Frame Examples

Scenario 1:

If an activity targets households and aims to see a change primarily in household level indicators, even if the activity targets a subset of beneficiaries with specific interventions, the baseline survey will need to develop one sampling frame. The sample size must be estimated for each key indicator and adjusted based on the proportion of households in the sub-set. The example as illustrated in Figure 1 describes a condition where one sampling frame will be adequate. Here, an activity targets 5,000 households for food assistance, and targets 3,000 households (a subset) with nutrition-specific interventions for women of reproductive age. The indicators to be measured are:

Figure 1

1. *Percent of households where women reported participating in decisions on the use of food assistance.*
2. *Percent of women of reproductive age consuming a diet of minimum diversity.*

For indicator 1, the sampling frame should be all households receiving food assistance. For indicator 2 the sampling frame is the 3,000 households with women targeted for the nutrition-specific intervention. In this scenario, there are two sampling frames.

[3]Adapted from "A Commissioner's Guide to Probability Sampling for Surveys at USAID", Julie Uwimana and Jennifer Kuzara (2020).

Box 3. Sampling Frame Examples, Continued

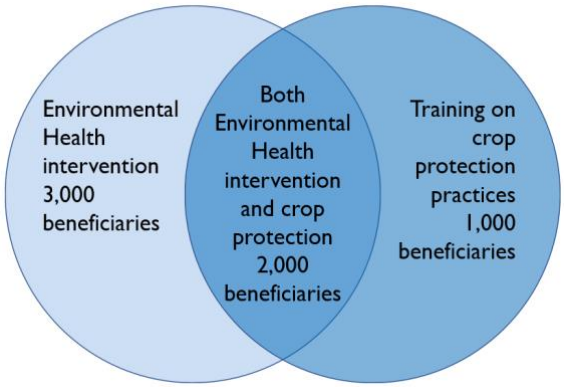
Scenario 2:

Assume, a multi-sectoral Agriculture and WASH activity (as illustrated in Figure 2) targets 5,000 beneficiaries to receive training on crop protection practices and 3,000 beneficiary households to receive a WASH intervention on solid waste disposal. Note that 2,000 beneficiaries who will receive crop protection training will also receive WASH intervention. However, 1,000 beneficiaries will only receive crop protection training. The activity will design a baseline/endline survey to measure two indicators:

1. *Percent of individuals who received training that are practicing appropriate crop protection procedures*
2. *Percent of households targeted by the WASH promotion activity that are properly disposing of solid waste.*

In this example, the baseline survey needs to develop three sampling frames - one with the 3,000 beneficiary households who receive only the WASH environmental health intervention (sampling frame 1), one with the 1,000 beneficiaries who will receive only crop protection training (sampling frame 2), and the third sampling frame with the 2,000 beneficiary households who will receive both WASH promotion activity and crop protection training(sampling frame 3).

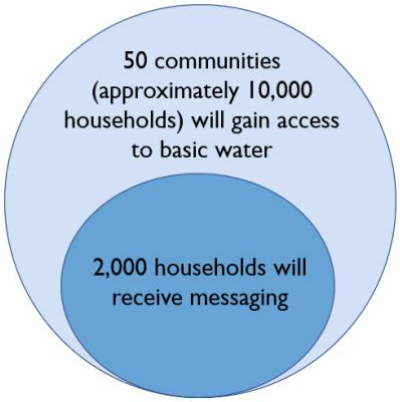
Figure 2



Scenario 3:

Assume a third scenario in which an activity targets 50 communities (approximate 10,000 households) to repair 20 water points to increase access to safe water. The activity also targets 2,000 households with messaging on hand washing at critical times. Figure 3 illustrates the example.

Figure 3



The activity will design a baseline/endline survey for the following two indicators:

1. *Percent of households using basic drinking water services*
2. *Percent of people targeted by the hygiene promotion program who know at least three (3) of the five (5) critical times to wash hands*

In this scenario, two sampling frames should be developed for indicator 1 in order to conduct a population-based baseline/endline two-stage cluster survey. For the first stage, a list of the 50 target communities is needed; and for the second stage, a list or count of all households within these communities is needed. For indicator 2, a third sampling frame will be needed for a beneficiary-based survey. The sampling frame should include all 2000 beneficiary households that received messaging on hand washing at critical times.

3.3 SAMPLING STRATEGIES

In this section we describe the different types of probability-based sampling strategies that can be used for population-based and beneficiary-based surveys. Non-probabilistic sampling methods[4] are not recommended when designing a survey because results are not generalizable to the entire target population. Probability-based sampling methods include simple random sampling (SRS), systematic random sampling, sampling with probability proportional to size (PPS), and multi-stage sampling.

3.3.1 ONE-STAGE SIMPLE RANDOM SAMPLING

If a list of all beneficiaries or households is available, and the logistical burden of data collection is reasonable, BHA recommends a one-stage simple random sampling (SRS) strategy.

A one-stage SRS design is advantageous because it is an equal probability of selection method and data is self-weighted which is necessary to generate unbiased estimates. Data collection in a SRS is simpler to implement and the resultant data is easier to analyze, reducing the chance of process and analytical errors. Analyzing data collected through a SRS design does not require advanced knowledge in survey statistics and sampling weights are not needed, making it ideal for emergency contexts where field teams prioritize timely implementation and immediate data over survey methodology.

Note: For a SRS, primary sampling units (direct beneficiaries) must be randomly selected from the sample frame, which should be the beneficiary register/database. In this approach, one cannot first select clusters (e.g., village, district, camps, and anything else but the primary sampling unit) and then select beneficiaries or households. The primary sampling units must be selected directly from the sampling frame. It is incorrect to estimate sample size using SRS in which the design effect is 1 (see Box 4), and then draw the sample using multiple stages.

To select a sample using SRS, first use a random number to generate a random number for each sampling unit in the sampling frame, then sort the sampling frame by this random number. If you are selecting a sample of 100 units, includ the first 100 units on the sorted sampling frame in the sample.

3.3.2 SYSTEMATIC RANDOM SAMPLING

For systematic random sampling, sampling units are ordered and selected according to a random starting point and fixed interval. This method is a variation of SRS and can be used when it is important to maintain the distribution of one or more attributes of the population in the selected sample or when a sample is taken while the sampling frame is being finalized. For example, a list of beneficiaries can be sorted by geographic region first and then systematically sampled to ensure that the sample represents all geographic regions in the same proportions as they are in the full population. Systematic random sampling can also be used when collecting baseline data during beneficiary registration.

To implement this sampling strategy, sort the sample by the attribute then randomly choose a starting point. Next, select every Nth sampling unit, where N is the sampling interval (SI). SI is determined by

[4] Non-probabilistic sampling methods use purposeful selection and judgement factors to choose sampling units so results cannot be extrapolated to the larger population from which the sample is selected.

dividing the total number of sampling units in the sampling frame by the desired sample size. Figure 2 below illustrates systematic random sampling. In this example, a total of 500 beneficiary households that received a WASH intervention make up the sampling frame. First, sort the sampling frame by region to ensure that sampled households will proportionately represent all regions in the target area. If you are selecting a sample of 100 households, the sampling interval is 500/100 or 5. Next generate a random number between one and the sampling interval to identify the starting point. In this case the 3rd sampling unit has been designated as the random starting point. Starting with household number 3, select every 5th household until the end of the sampling frame is reached.

Figure 2. Systematic Random Sampling

Household Number	Region	Name	Activity
1	Region 1	Household Name	WASH
2	Region 1	Household Name	WASH
3	Region 1	Household Name	WASH
4	Region 1	Household Name	WASH
5	Region 1	Household Name	WASH
6	Region 1	Household Name	WASH
7	Region 2	Household Name	WASH
8	Region 2	Household Name	WASH
9	Region 2	Household Name	WASH
10	Region 2	Household Name	WASH
11	Region 2	Household Name	WASH
12	Region 2	Household Name	WASH
13	Region 3	Household Name	WASH
14	Region 3	Household Name	WASH
15	Region 3	Household Name	WASH

3.3.3 SAMPLING WITH PROBABILITY PROPORTIONAL TO SIZE (PPS)

The PPS method ensures that villages with more households have a greater chance of being selected compared to villages with fewer households, thus giving each household an equal likelihood of being selected at the second stage. The probability proportional to size (PPS) method of sampling is commonly used in surveys when selecting villages or communities as part of a two-stage cluster sample (see below).

In order to use PPS sampling, the partner must have accurate information on the total size of each cluster (e.g. village population for a population-based survey or the number of individual beneficiaries or beneficiary households targeted for a beneficiary-based survey). When analyzing data generated from a two-stage cluster sample with PPS sampling at the first stage and an equal number of sampling units being selected at the second stage, weighting is not necessary since all sampling units have the same probability of selection.

Suppose you are selecting a sample of households that receive WASH interventions for a WASH beneficiary survey using a two-stage cluster sample. The following steps are used to select the first-stage sample of villages using PPS:

1. Construct the sampling frame by listing all villages where beneficiary households are located along with the number of beneficiary households in each village.
2. Calculate the cumulative total of households.
3. Divide the overall cumulative number of households by the number of villages to be sampled to determine the sampling interval (SI).
4. Generate a random number between one and the SI to determine which village to sample first.
5. Identify which village contains the random start household; this will be the first sampled village.
6. Add the sampling interval to the random start value to select the next village; so on and so forth, until the desired number of villages have been selected.

This process is demonstrated in Figure 3 below: Five villages were selected from 20 villages using a sampling interval (SI) of 2,500/5=500, where 2,500 is the total number of households in the area (cumulative household total). A random start (RS), 227, was generated using a random number generator (bound by 1 and 500). Village number 4 contains the 227th household so it is the first sampled village. The remaining villages are sampled by incrementally increasing the RS by the SI (RS, RS+SI, RS+2SI, RS+3SI, RS+4SI).

Figure 3. Sample Selection Process

Village	No. of Households	Cumulative Households	Sample Selection
Village 1	50	50	
Village 2	100	150	
Village 3	50	200	
Village 4	300	500	RS = 227
Village 5	50	550	
Village 6	100	650	
Village 7	200	850	227+500=727
Village 8	50	900	
Village 9	150	1050	
Village 10	50	1100	
Village 11	50	1150	
Village 12	100	1250	727+500=1227
Village 13	50	1300	
Village 14	450	1750	1227+500=1727
Village 15	200	1950	
Village 16	100	2050	
Village 17	150	2200	
Village 18	50	2250	1727+500=2227
Village 19	150	2400	
Village 20	100	2500	

3.3.4 MULTI-STAGE SAMPLING

Multi-stage sampling is often preferred for its functionality and cost-effectiveness compared to simple random sampling. Separate sampling frames are constructed for each stage of sampling. For example, the first stage sampling frame might include all villages in the target area, the second stage sampling frame might include all households within these villages and the third stage sampling frame might include all individuals within these households. The sample is designed to select units at the lowest level of sampling. Multistage sampling will commonly use multiple approaches to sampling within the various stages.

Two-stage cluster sampling is a special kind of multi-stage sampling where the target population is first divided into clusters; these clusters are sampled and then a second sample is selected from each of the sampled clusters. Two-stage cluster sampling designs are typically used in surveys when the logistical costs of data collection using a one-stage SRS are high because the communities in the target population are too far apart and the budget prohibits data collectors to travel to all areas in the target population. This strategy is also suitable when a list of all participants is not available from which to develop a sampling frame of direct beneficiaries. A cluster design can be a cost-efficient way to sample a geographically dispersed population.

In a two-stage cluster sampling design, the first stage involves randomly selecting clusters (i.e. villages/ communities/ groups) from a list of all clusters. In the second stage, households or individuals are randomly selected from the sampled clusters.

While cluster sampling may be more cost-effective, the approach provides less precision than SRS. Households within a cluster (e.g. village) tend to be more similar to each other than to households in other clusters, which is known as intracluster correlation. To minimize intra-cluster correlation, BHA recommends partner sample more clusters with a smaller sample from each cluster. For example, any of the following options can be used to collect data from 660 sampling units.

1) 22 clusters x 30 sampling units = 660
2) 33 clusters x 20 sampling units = 660
3) 44 clusters x 15 sampling units = 660

The logistical burden will likely be lighter for option 1, compared to option 3. Using option 3 is preferable and will increase the power but it may also increase the logistical burden and cost. Therefore, partners should carefully consider the cost and advantage to determine the sampling options.

Figure 4 below provides a simplified example whereby Villages 1 and 2 (out of 4 total villages) were sampled in the first stage of selection using PPS as described in Section 3.3.3. Then in the second stage of the selection, three beneficiary households were selected from each village using SRS. Using a random number generator from 101 to 107 (for village 1), households 102, 104, and 105 were selected. Using a random number generator from 201 to 215, households 202, 208, and 211 were selected from Village 2.

Figure 4. Multi-stage Sampling Process

Village No.	Village Name	Household Number	Head of Household	Activity
1	Village 1	101	Sample Name	WASH
1	Village 1	102	Sample Name	WASH
1	Village 1	103	Sample Name	WASH
1	Village 1	104	Sample Name	WASH
1	Village 1	105	Sample Name	WASH
1	Village 1	106	Sample Name	WASH
1	Village 1	107	Sample Name	WASH
2	Village 2	201	Sample Name	WASH
2	Village 2	202	Sample Name	WASH
2	Village 2	203	Sample Name	WASH
2	Village 2	204	Sample Name	WASH
2	Village 2	205	Sample Name	WASH
2	Village 2	206	Sample Name	WASH
2	Village 2	207	Sample Name	WASH
2	Village 2	208	Sample Name	WASH
2	Village 2	209	Sample Name	WASH
2	Village 2	210	Sample Name	WASH
2	Village 2	211	Sample Name	WASH
2	Village 2	212	Sample Name	WASH
2	Village 2	213	Sample Name	WASH
2	Village 2	214	Sample Name	WASH
2	Village 2	215	Sample Name	WASH

3.3.5 STRATIFIED SAMPLING

Some partners may consider using stratified sampling; however, it is generally discouraged by BHA due to its complexity in weighting and analyzing the data. **It is crucial that sampling weights are used to produce accurate estimates when using stratified sampling.** In stratified samples, the sampling frame is divided into homogenous groups, i.e., those with similar characteristics. These groups are referred to as "strata." A sample is then drawn randomly from each stratum. Common characteristics (or variables) used for stratification are geographic regions, sex categories, and intervention activity type.

Stratified sampling designs are typically used in one of two instances: (1) when the outcome of interest is strongly correlated with the characteristics (variables) that were used for the stratification. For example, if beneficiary households received different intervention types and the partner wished to stratify by intervention type, the sampling frame could be separated into separate groups based on the type of

intervention received. (2) Stratified sampling may also be used to ensure under-represented groups (who may not be represented using random selection methods). In this latter case, oversampling may be used to gather a disproportionate number of sampling units (e.g. households, individuals) from a strata of interest. This approach ensures that the sample will have sufficient data to support sub-analyses of characteristics that are typically low prevalence at the population level.[5]

3.4 SAMPLE SIZE CALCULATION

After determining the type of survey and sampling strategy, it is important to determine how many individuals or households (called "sampling units") will be needed in order to generate an accurate estimate for the indicators being measured. In this context, the sampling estimates are values we generate from the survey sample that we use to make a best guess about the true (but not directly observable or knowable) values within the population or group of beneficiaries. For example, if we want to know what percent of beneficiary households are currently using promoted handwashing practices, we would need to know how many households to interview in order to generate an accurate estimate of what all beneficiary households were generally doing in terms of handwashing.

The three critical pieces of information needed to determine the appropriate formula for estimating the sample size are:

1. the purpose of the survey, either comparative (baseline/endline) or descriptive, point estimate (monitoring);
2. the indicator(s) the survey data will be used to estimate and how they are expressed, i.e. proportion, mean or total - see Table 5; and
3. the sampling strategy that is being used, either one-stage SRS or two-stage cluster sampling.

Table 5. Examples of Indicators Expressed as Proportions, Means, and Totals

Indicator Expression	Indicator
Proportion	Percent of households practicing handwashing Percent of households with poor FCS score
Mean	Average (mean) yield for targeted agricultural commodity Average (mean) rCSI score
Total	Number of hectares under improved management practices

NOTE: Since the majority of outcome indicators used in BHA surveys are expressed as either proportions or means, this abbreviated guidance does not provide guidance for calculating sample size for totals which are rarely used and not typically recommended. If an applicant identifies key outcome indicators expressed as a total for a survey, please contact the BHA M&E Advisor responsible to provide backstopping support to the award.

In the next two sections, we provide guidance on how to calculate a sample size for baseline/endline surveys and for monitoring surveys. Box 4 describes key parameters that are used as inputs for the sample size calculations provided.

[5] Adapted from "A Commissioner's Guide to Probability Sampling for Surveys at USAID", Julie Uwimana and Jennifer Kuzara (2020).

Box 4. Key Terms Used in This Section for Sample Size Calculations

- **Estimated proportion or mean:** This is the survey estimate of the true (but unknown) population proportion or mean at the time of the survey.
- **Standard deviation:** The standard deviation is a measure of dispersion in the sample distribution for an indicator, and is expressed in the same units as the indicator.
- **Critical value of normal probability distribution (z-value):** The point on the normal probability distribution curve that corresponds to a specific level of confidence in the sample estimate. We typically use a 95% confidence level. The z-value for a 95 percent confidence level is 1.96 for a two-sided test and 1.64 for a one-sided test.
- **Effect Size:** The effect size is the targeted amount of change to be measured when comparing two data points, i.e. from baseline to endline. The smaller the amount of change to be measured, the larger the sample size.
- **Margin of Error:** The margin of error is the amount of error considered to be acceptable in estimating the proportion or mean. This value is typically set between 5 and 10 percent. The larger the acceptable margin of error, the smaller the sample size.
- **Design Effect:** The design effect helps to measure the sampling error associated with how the sampling was designed and carried out. In two-stage cluster designs where households are selected after communities were selected, we use a design effect of 2 as a rule of thumb, unless a more accurate estimate of the design effect can be made based on previous or similar survey data. The design effect of 2 indicates that the sampling error is twice that compared to using a single-stage SRS design.
- **Non-response:** In surveys, it is expected that some people who are selected to participate in the survey will not be available or may not be willing to complete the survey. This is called non-response, and must be taken into account when calculating the sample size. We can use a nonresponse rate of 10% as a rule of thumb until a more accurate estimate is available (based on previous survey data).

3.4.1 CALCULATING SAMPLE SIZE FOR BASELINE/ENDLINE SURVEYS

For baseline/endline surveys we need to calculate the appropriate sample size for **comparing the values of indicators collected at two points** in time: at the start of the activity and after the activity is completed. In order to do this, we need to know whether we are collecting data using a one-stage random sample or a two-stage cluster sample and we need to know what type of indicator the sample is being designed for.

The below two examples show the sample size calculations for a baseline indicator expressed as a proportion and a baseline indicator expressed as a mean using both a one-stage SRS sampling strategy and a two-stage cluster sampling strategy.

In Example 1, we use "Percent of households with poor FCS score" to estimate sample size, This sample size calculation is relevant for comparing any indicator expressed as a proportion at two points in time. The parameters used are: 1) baseline proportion of 50 percent, 2) expected endline proportion of 40

percent (effect size of 10 percentage points), and 3) an expected non-response rate of 10 percent. The resulting sample size is 339 for an SRS strategy and 678 for a two-stage cluster sampling strategy. Note that the only difference between calculating the sample size for a single-stage SRS compared to a two-stage cluster sample is an increase in the design effect from one to two.

BHA recommends using these sample sizes for comparing indicators expressed as proportions unless the partner has more reliable information on the estimated baseline proportion or the expected non-response rate; or if the partner is setting a target for endline other than a 10 percentage point change. Note that the targeted percentage point change is the main driver for determining the sample size. An increase in the targeted percentage point change will result in a smaller sample size; likewise a decrease in the targeted percentage point change will result in a larger sample size.

Example 1. Calculating Sample Size for Comparing Indicators (between Baseline and Endline) Expressed as a Proportion

INDICATOR: Percent of households with poor FCS score		
	Single stage SRS	**Two stage cluster sample**
Estimated baseline proportion	50% (0.5)	50% (0.5)
Expected endline proportion	40% (0.4)	40% (0.4)
Effect size (expected change)	10 percentage points	10 percentage points
Confidence level (one-sided z-value)	95% (1.64)	95% (1.64)
Power level (z-value)	80% (0.84)	80% (0.84)
Design effect	1	2
Initial sample size	305	610
Expected level of non-response	10%	10%
Final sample size	339	678

To calculate sample size for indicators with parameters other than those shown above, a sample size calculator can be used; such as that provided by Feed the Future as part of their Population-Based Sampling Guidance.[6]

NOTE: Use tab 1 of the Feed the Future Sample Size Calculator available on the USAID website for computing baseline/endline sample size when comparing indicators expressed as proportions.

For endline surveys, information regarding the baseline proportion, design effect and nonresponse rates should be available and used for adjusting the endline sample size. See section below regarding adjustments for sample size at endline.

In Example 2, we provide a sample size calculation for comparing an indicator expressed as a mean at two points in time. Here, the Activity is providing seed inputs to farmers and expecting the average yield

[6] Refer to "Feed the Future Population-Based Survey Sampling Guide and Calculator", USAID/FANTA/FHI360 (2018).

in millet across all farmers over the life of the activity to increase by .10 metric tons from baseline to endline. For this calculation, the estimated baseline/endline mean value and standard deviation are needed. The calculation yields a sample size of 220 for an SRS strategy and 440 for a two-stage cluster sampling strategy. Since yield for targeted agricultural commodities are reported by individual farmers, the Activity would need a list of all farmers receiving seeds to use an SRS strategy to select the 220 farmers.

Example 2. Calculating Sample Size for Comparing Indicators (between Baseline and Endline) Expressed as a Mean

INDICATOR: Yield of targeted agricultural commodity (millet)		
	Single stage SRS	**Two stage cluster sample**
Estimated baseline mean	1.30 MT/Ha	1.30 MT/Ha
Expected endline mean	1.40 metric MT/Ha	1.40 MT/Ha
Effect size (expected change)	0.10 MT/Ha	0.10 MT/Ha
Estimated standard deviation of the baseline mean	0.40 MT/Ha	0.40 MT/Ha
Estimated standard deviation of the endline mean	0.40 MT/Ha	0.40 MT/Ha
Confidence level (one-sided z-value)	95% (1.64)	95% (1.64)
Power level (z-value)	80% (0.84)	80% (0.84)
Design effect	1	2
Initial sample size	198	396
Non-response adjustment	10%	10%
Final sample size	220	440

NOTE: Use tab 2 of the Feed the Future Sample Size Calculator to calculate baseline/endline sample size when comparing indicators expressed as means.

Inflating for the Number of Households to Contact

When conducting a population-based survey that is designed to study a target population of individuals, rather than households, there may be a need to inflate the sample size of households to account for households that may not include an eligible member of the target population. For example, if we are designing the sample to detect a change in exclusive breastfeeding for children under 6 months and we need 200 children, then we will likely need to contact more than 200 households since all households may not include a child under 6 months. In this case an inflation adjustment should be made.[7]

[7] See "Feed the Future Population-Based Sampling Guide", Section 2.2.4, pages 22-23, USAID/FANTA/FHI360 (2018). The Feed the Future Sample Size Calculator can be used to apply this inflation adjustment.

Inflating for Nonresponse

BHA recommends using an expected nonresponse rate of 10 percent for sample size calculations. However, in some cases the nonresponse adjustment should be increased, for example in cases where the contact information on the sampling frame may not be reliable or where it may be difficult to access households or individuals due to security or other reasons.

Adjusting the Sample Size at Endline

The baseline/endline sample size calculated in the examples above are assumed to be the same. However, there are instances when the endline sample size should be re-calculated and adjusted if needed. After the baseline survey is completed, the actual number of households or individuals interviewed will be known. This number may fall short of the desired sample size due to higher than expected nonresponse or some other reason. In this case the endline sample size may need to be increased to compensate for the shortage at baseline. Since the parameters which were estimated at baseline (baseline proportion, design effect and nonresponse level) can now be calculated, the endline sample size should be recalculated taking these actual values into account.[8]

NOTE: Use tab 5 (for comparing proportions) and tab 6 (for comparing means) of the Feed the Future Sample Size Calculator for computing the adjusted sample size at endline.

Computing the Sample Size when Using Multiple Indicators

If a survey is being designed to collect data on multiple indicators, then it will be important for the partner to calculate the appropriate sample size for each indicator. Once the sample sizes are known for each indicator, the maximum of all of the computed sample sizes should be used.

3.4.2 CALCULATING SAMPLE SIZE FOR MONITORING SURVEYS

Most BHA emergency indicators are outputs collected through non-survey routine monitoring methods. The list of BHA emergency indicators also include outcome indicators. Sometimes partners collect data for outcome indicators with the post distribution monitoring surveys. If a partner collects indicator data through a post distribution monitoring (PDM) survey, the design should use a probabilistic sampling method. **Partners should reference each indicator PIRS for guidance to determine whether representative surveys are appropriate for routine monitoring**.

For monitoring surveys, partners should calculate the appropriate sample size for estimating an indicator that takes into consideration the **purpose** of the data collection:

- To verify outputs, quality, and process monitoring: If the purpose of the monitoring data is to conduct ongoing process monitoring, or verify that distributions have been received to the expected level of quality (e.g. post-distribution monitoring following each distribution) then the **sample size should be calculated using the sample size calculation for estimating an**

[8] See "Feed the Future Population-Based Sampling Guide", Section 2.2.7, pp. 26-29, USAID/FANTA/FHI360 (2018).

indicator at one point in time. Partners may use a wider margin of error (up to +/- 10 percent) for monitoring surveys.

- Similar to baseline/endline surveys, we need to know whether we are collecting data using a one-stage random sample or a two-stage cluster sample and we need to know what type of indicator the sample is being designed for. We provide two examples below that demonstrate sample size calculations for an indicator at one point in time expressed as a proportion and a mean using both a one-stage SRS sampling strategy and a two-stage cluster sampling strategy. If an indicator will not be used to compare to baseline and endline, a higher margin of error (up to +/- .10) may be appropriate to keep the sample size manageable.
- For longer awards, compare mid-line indicator values with baseline and endline: If the purpose of the monitoring data is to measure outcome indicators that will be **statistically compared to baseline and endline** values or at different points in time, the sample size should be calculated using the sample size calculation for comparing two indicator values as described in Section 3.4.1 above.
- If a survey has dual purposes, e.g., statistical comparison between baseline and endline and to conduct basic verification and process monitoring - then the sample size should be calculated for both purposes, and the higher of the two sample sizes should be used.

Example 3 illustrates sample size calculations for a point estimate of an indicator expressed as a proportion. In this example, the Activity provided shelter and settlement non-food based items (NFIs) to 2,000 beneficiary households and would like to know the percentage of beneficiary households that report being satisfied with the quality of the NFIs received. The expected proportion of satisfied beneficiaries is 80 percent and the acceptable margin of error is plus or minus 6 percent. The single-stage SRS should be used when a list of beneficiary households is available and in this case the sample size would be 297 households. This survey would require a sample size of 594 households. Note that the acceptable margin of error is the main driver for estimating the sample size. Increasing the margin of error will result in a smaller sample size while decreasing the margin of error will result in a larger sample size. BHA recommends using a margin of error between +/- 5 percent and +/-10 percent.

Example 3. Calculating Sample Size for Point Estimates of Indicators Expressed as a Proportion

INDICATOR: S7: Percent of beneficiaries reporting satisfaction with the quality of the NFIs received Shelter & Settlements Non-food items (NFIs)		
	Single stage SRS	**Two stage cluster sample**
Expected proportion at time of survey	80% (0.5)	80% (0.5)
Margin of error*	+/- 6%	+/- 6%
Confidence level (two-sided z-value)	95% (1.96)	95% (1.96)
Design effect	1	2
Initial sample size	171	342
Expected level of non-response	10%	10%
Final sample size	190	380

*The hypothetical partner in this example is choosing to use a margin of error of +/- 6 percent, which is within the range of 5 to 10 percent MOE recommended by BHA.

NOTE: Use tab 3 of the Feed the Future Sample Size Calculator to calculate sample size for point estimates of indicators expressed as proportions.

Example 4 shows sample size calculations for a point estimate of an indicator expressed as a mean. In this example, the Activity is estimating the sample size needed to generate a point estimate of the household food consumption score. The estimated mean at the time of the monitoring survey is 30 with a standard deviation of 14. The acceptable margin of error for the estimate is plus or minus eight percent. Eight percent of the mean is .08 times 30 or 2.4. With a sample size of 146 and using a single-stage SRS strategy, the point estimate of 30 will reflect a true mean between 27.6 and 32.4. Raising the acceptable margin of error will result in a smaller sample size.

Example 4. Calculating Sample Size for Point Estimates of Indicators Expressed as a Mean

INDICATOR: Food Consumption Score (FCS)		
	Single stage SRS	**Two stage cluster sample**
Estimated mean at the time of the monitoring survey	30	30
Estimated standard deviation of the mean at the time of the monitoring survey	14	14
Margin of error	+/-8% or 2.4	+/-8% or 2.4
Confidence level (two-sided z-value)	95% (1.96)	95% (1.96)
Design effect	1	2
Initial sample size	131	262
Expected level of non-response	10%	10%
Final sample size	146	292

*The hypothetical partner in this example is choosing to use a margin of error of +/- 8 percent, which is within the range of 5 to 10 percent MOE recommended by BHA.

NOTE: Use tab 4 of the Feed the Future Sample Size Calculator to calculate sample size for point estimates of indicators expressed as means.

3.5 SAMPLE WEIGHTING AND DATA ANALYSIS FOR PROBABILITY SURVEYS

For quantitative surveys, the partner should describe how the baseline and endline data will be statistically compared, as appropriate. Endline survey data should be analyzed and compared with baseline data as part of the final evaluation, including statistical tests of differences in key outcome indicators. For all indicators listed in table A, detecting change(s) requires using a statistical package (i.e SPSS, Stata, SAS, CSPro, or other statistical application) and conducting a test of difference. For FCS, partners should test the difference between baseline and endline FCS raw score as well as "% of households with acceptable FCS score". For HHS, partners should test the difference between baseline

and endline HHS raw score as well as "% of households with moderate or severe Household Hunger Scale (HHS) score".

The baseline and endline surveys should follow the same methodology to simplify analysis. While partners can take a census at baseline and a sample at endline, BHA encourages that a sample be taken at both baseline and endline for efficiency. When testing for differences between baseline and endline values, it is important to use the appropriate statistical tests for the type of data being analyzed.[9] Partners may consult with their AOR and the BHA M&E Advisor backstopping their country to discuss their plan.

Table 6. Indicator and Testing Method

Indicator	Indicator title	Test
Food Consumption Score	FCS raw score	Two-sample t-test; One-sample t-test be used when the baseline data was collected through census
	% of households with acceptable FCS score	Pearson's chi-squared test
Household Hunger Scale	HHS mean score	Two-sample t-test; One-sample t-test be used when the baseline data was collected through census
	% of households with moderate or severe HHS score	Pearson's chi-squared test
Reduced Coping Strategy Index	rCSI mean score	Two-sample t-test; One-sample t-test be used when the baseline data was collected through census
Knowledge of Critical Moments for Handwashing	Percent of beneficiaries who know 3-5 critical moments for handwashing	Pearson chi-squared test
Satisfaction with Shelter and Settlement Non-food Items (NFIs)	Percent of beneficiaries reporting satisfaction with the quality of the NFIs received - Shelter & Settlements NFIs	Pearson chi-squared test

[9] For example, when a baseline uses a census of participants and the endline uses a survey, a one-sample t-test may be used to test for differences in continuous variables. When a survey is conducted at both baseline and endline, then a two-sample t-test may be used.

CHAPTER 4: MONITORING

Monitoring plays an important role in ensuring that emergency activities are efficient, effective, and on track to meet their objectives. Monitoring should enable partners to track progress, ensure accountability, and adaptively manage their awards. The Monitoring Approach submitted at Application should be based on the activity's planned interventions and anticipated results and be designed to facilitate timely management decisions. A well-designed monitoring system can provide credible and actionable data enabling both the partner and BHA to gain important insights into how to manage and improve the effectiveness of the activity.

Examples of the role that monitoring plays for accountability and performance include:

- Demonstrating results to stakeholders
- Accountability to the affected population
- Keeping to the activity plan during implementation
- Improving the relevance and appropriateness of the activity
- Identifying implementation issues and improving the quality of implementation
- Organizational learning to inform future activities

This chapter provides BHA's guidance for both performance monitoring and context monitoring in emergency activities. **Performance monitoring** includes monitoring the quantity, quality, and timeliness of activity outputs within the control of BHA partners, as well as monitoring activity strategic outcomes that are expected to result from the combination of these outputs and other factors. **Context monitoring** includes monitoring local conditions or external factors that are outside of the manageable interests of the partner but may directly affect implementation and performance.

4.1 PERFORMANCE MONITORING

Performance monitoring is defined as the ongoing and systematic collection of performance indicator data and other quantitative or qualitative information to reveal whether implementation is on track, the quality of implementation is high, and whether expected results are being achieved. This includes monitoring the quantity, quality, and timeliness of activity outputs within the control of BHA or its partners, as well as the monitoring of activity and strategic outcomes that are expected to result from the combination of these outputs and other factors.[10] In the context of BHA emergency awards, performance monitoring falls into three distinct groups: outcome monitoring, output monitoring, and process monitoring.

4.1.1 OUTPUT MONITORING

Outputs are the immediate products of interventions implemented by an activity, including goods and/or services provided (e.g., food or cash distributed), training completed, and behavior change communication events held. Outputs are what are produced as a direct result of inputs. They are the tangible, immediate, and intended products or consequences of an activity within BHA/partners' control or influence. Outputs must be completed in order for an activity to achieve its outcomes.

[10] Adapted from ADS 201.3.5.5A, *ADS Chapter 201 Program Cycle Operational Policy*, USAID (2017).

Monitoring outputs is a critical tool for both project management and accountability. It allows the stakeholders to understand whether the implementation is on track as planned, and whether it corresponds to the resources spent. Output monitoring is typically conducted through routine monitoring approaches. Routine monitoring refers to data that is collected on an ongoing basis by activity staff throughout implementation.

Data Collection for Output Monitoring: Partners most often use routine monitoring methods to track progress on outputs. This may include approaches such as using checklists or other tools to track distributions, number of beneficiaries trained, or other outputs.

4.1.2 OUTCOME MONITORING

Outcomes are the conditions of people, systems, or institutions that indicate progress or lack of progress toward achievement of activity goals. Outcomes are any result higher than an output to which a given output contributes but for which it is not solely responsible. Outcomes may be intermediate or end outcomes, short-term or long-term, intended or unintended, positive or negative, direct or indirect[11] These might include changes in households' food security or nutrition status, or changes in people's knowledge, attitudes, or practices.

Monitoring outcomes is important to understand if an activity is achieving or on-track to achieve the stated Purposes, Sub-Purposes, Intermediate Outcomes, and Outcomes (as applicable). Partners should document their outcome monitoring strategy in their Monitoring Approach, including what methods and practices will be used to monitor outcomes and the frequency of data collection.

Outcome monitoring can be particularly challenging in the context of rapid onset emergencies. Emergency activities are typically implemented over a short period of time (12 months or fewer) limiting the ability to measure changes in some indicators. The affected populations may be mobile (refugees and internally displaced persons (IDPs)), which can make it difficult to re-sample the same population. Finally, security issues can limit access to the affected population. This section describes some methods and practices that can be used for outcome monitoring in these environments.

Data Collection for Outcome Monitoring: Partners use a variety of methods to collect information for outcome monitoring. This may include quantitative methods such as routine monitoring methods (e.g. checklists); beneficiary-based surveys (e.g. post-distribution monitoring surveys); or population-based surveys. Partners may also use qualitative methods such as key informant interviews and focus groups to help inform outcome monitoring. This may involve direct data collection and/or remote data collection. *Note that BHA does not recommend using a census to collect outcome monitoring data.*

4.1.3 PROCESS MONITORING

Process monitoring allows activity managers to assess implementation quality, adherence to minimum standards, and identify ways in which implementation can be improved. Process monitoring is a critical

[11] Adapted from ADS 201.6 definitions. *ADS Chapter 201 Program Cycle Operational Policy*, USAID (2017).

tool for managers as it allows for early detection of issues. Identifying and addressing implementation issues early is important so that outputs are of high quality and activity objectives are likely to be achieved.

With resource transfer interventions (regardless of modality), the aim of process monitoring is to observe implementation quality, ensure accountability across the supply chain and ensure that the participant experience throughout the program cycle meets or exceeds humanitarian standards. Process monitoring may be used to identify protection issues including accommodation for vulnerable groups, sexual exploitation and abuse, and transaction costs incurred, or issues of fraud, waste, or abuse.

Process monitoring for supply chains should ensure that tracking systems and standard operating procedures are able to effectively follow resource transfers to the end recipient. This includes methods for monitoring and minimizing losses (including adequate storage, transportation, and handling), ensuring commodity quality, and adherence to checks and balances which specifically assign responsibility.

The objective of process monitoring is to ensure that activities and resources are delivered in such a way that it meets or exceeds humanitarian standards. More specifically, this includes timely delivery of appropriate assistance while doing no harm in the process and minimizing exposure to risks (with specific consideration to protection and gender).

Process monitoring can be used to identify:

- Whether assistance was received by the right person, safely, on time, and in the correct amount
- If travel and wait times to receive assistance are appropriate
- Whether any transaction costs were incurred in receiving the assistance
- Targeting-related inclusion and exclusion errors
- Quality of training sessions, and social and behavior change sessions
- Quality of demonstration plots, or inputs provided by an activity
- Areas for to improve the implementation quality

One example of process monitoring is food basket monitoring. The purpose of food basket monitoring is to ensure consistency in the size of the ration participants are receiving. For example, a sample of participants leaving the distribution site might have their ration weighed to ensure that it is within the margin of error of the planned ration size. (In this particular instance it may be appropriate to use LQAS to determine the sample size since the objective of the data collection is to determine whether the quality of the ration item being tested is either above or below a certain predetermined standard/threshold for minimum acceptable quality.)

Data Collection for Process Monitoring: Partners use a variety of methods to collect information for process monitoring. This may include qualitative methods and tools such as observation, interviews, and group discussions; quantitative methods such as routine monitoring methods (e.g. checklists); beneficiary-based surveys (e.g. post-distribution monitoring surveys). This may involve direct data collection and/or remote data collection. *Note that BHA does not recommend using a census to collect process monitoring data.*

For additional guidance on process monitoring, see the Monitoring Guidance for CTP in Emergencies, Cash Learning Partnership.

4.2 CONTEXT MONITORING

In addition to monitoring the performance of an activity, BHA recommends partners to also monitor the surrounding context. Context monitoring is defined as the systematic collection of information about conditions and external factors relevant to the implementation and performance of an activity. This includes information about local conditions that may directly affect implementation and performance (such as other activities operating in the same sector or geographic area), markets, conflicts, seasonal natural hazards, or external factors that may indirectly affect implementation and performance (such as macroeconomic, social, security or political conditions). Context monitoring should be used to monitor assumptions and risks identified in an activity's ITT.[12]

Applicants must include context monitoring as a section of their Monitoring Approach that describes the operational context issues that may impact the activity and how these issues will be monitored. This section must identify the indicators and data collection methods that will be used.

Data Collection for Context Monitoring: Partners use a variety of methods to collect primary and secondary data for context monitoring. This may involve direct data collection and/or remote data collection.

For activities using cash and voucher assistance (CVA) or in-kind food to achieve food security outcomes,: the operational context monitoring plan should monitor the price and/or availability of staple food commodities in the market areas[13] where operations are occurring. When appropriate, applicants should identify the commodities that will be tracked, the locations, and the frequency of market monitoring. To mitigate duplicative monitoring, partners may use reliable secondary data from other actors (UN, FEWS NET, NGOs, and/or National Ministries) in shared markets. If specific thresholds are to be established to signal the possibility of a distortion, describe the process that will be used to identify those thresholds. BHA encourages partners to work with FEWS NET, WFP, Food Security Clusters and Cash Working Groups to ensure harmonized technical standards around market monitoring including units of measurement (both in terms of weight and commodity specifications), frequency of collection, methodology and locations.

For more technical guidance on market monitoring and analysis see the following resources:

- MARKit: Price Monitoring, Analysis and Response Kit , Catholic Relief Services (CRS) (2015).
- ICRC Market Analysis Guidance: Chapter 3, International Federation of Red Cross and Red Crescent Societies (ICRC) (2014).
- WFP Price Monitoring, WFP (2017).

[12] Context monitoring definition adapted from ADS 201.3.5.5b. *ADS Chapter 201 Program Cycle Operational Policy*, USAID (2017).

[13] For potential marketplaces to monitor, please see typology and guidance from the MarKIT tool p.19-28

4.3 MONITORING APPROACHES

4.3.1 POST DISTRIBUTION MONITORING (PDM)

PDM is a performance and process monitoring tool primarily used to monitor the use and quality of transfers (in-kind, cash, and voucher), wait time, distance to distribution centers, effectiveness of the complaints and feedback mechanism, and other factors such as taste of food, content, quality and quality of NFIs, and adequacy of the distribution. PDM provides managers with information which they can use to assess the appropriateness of the modality, the efficiency of implementation, and the effectiveness of the approach to achieve stated outcomes.

PDM often tracks utilization of household food or non-food assistance, timeliness of the assistance, participants' perception about gender and protection considerations, safety and security, access to and effectiveness of participant feedback loops and other factors associated with the transfer of the entitlement. The frequency of PDM depends on the design of the activity. Justification for the proposed frequency should be clearly communicated in the PDM section of the Monitoring Approach.

The PDM approaches proposed for the activity must be documented in the PDM section of the Monitoring Approach. It must include the following components: indicators collected, survey design, sampling frame, sample size calculation, sample selection, and analysis. PDM data can be collected through routine monitoring or through surveys. If it is collected through a survey, the design should use a probabilistic sampling method. However, considering the frequency of PDM and the purpose, a lower level of precision could be acceptable to keep the sample size at a reasonable size. Note that sampling weights are not necessary if a simple random sample (SRS) or probability proportional to size (PPS) sampling method is used; it is beneficial, however, to weight the data if a two-stage cluster design is used. BHA recommends that applicants include sample size calculations for PDMs in the M&E Plan at the application stage; partners should update these calculations, as needed, post-award.

4.3.2 BENEFICIARY FEEDBACK AND COMPLAINT MONITORING AND AAP

Participant complaint and feedback monitoring is both an important performance monitoring tool and is necessary for operationalizing accountability to affected populations (AAP). In line with the fourth and fifth Core Humanitarian Standards, the Monitoring Approach must describe the beneficiary complaint and feedback monitoring system and how the data is used for adaptive management. This includes:

- How the affected population will be made aware of the complaint and feedback mechanism
- An overview of the participant complaints and feedback mechanism including a description of:
    - the proactive and reactive channels that will be put in place to receive complaints and feedback from participants (e.g., hotline, suggestion box, focus group discussions)
    - how face-to-face complaints and feedback are documented
    - the referral pathways
    - the complaint and feedback categories
    - the feedback loop closure verification, satisfaction and documentation process
- How the complaint and feedback mechanism is routinely tested for functionality
- How the participant complaint and feedback mechanism appropriately covers the implementation area, especially the most marginalized and hardest to reach

- Indicators and targets that will be used to track the level of timeliness, quality and satisfaction of the resolution of complaints and the level of adaptive management that results from the resolution of complaints and feedback
- How the AAP data is reported (including demographics, analysis of trends, and summary of challenges and adaptations), its frequency and key audience
- How the participant complaint and feedback mechanism will be appropriately resourced (staffing and budget)

BHA expects that the complaint and feedback mechanism will be accessible and inclusive to beneficiaries, the existence of the mechanism will be well known among the participant population and that the feedback loop will be closed.

Box 5. Humanitarian Standards

The 4th and 5th Core Humanitarian Standards

1. Humanitarian response is appropriate and relevant.
2. Humanitarian response is effective and timely.
3. Humanitarian response strengthens local capacities and avoids negative effects.
4. **The humanitarian response is based on communication, participation and feedback.**
5. **Communities and people affected by crisis should have access to safe and responsive mechanisms to handle feedback and complaints).**
6. Humanitarian response is coordinated and complementary.
7. Humanitarian actors continuously learn and improve.
8. Staff are supported to do their job effectively, and are treated fairly and equitably.
9. Resources are managed and used responsibly for their intended purpose.

4.3.3 REMOTE MONITORING

BHA partners work in many complex non-permissive environments (NPEs) where security concerns prevent staff from conducting regular site visits to monitor and verify the implementation of activities and results. USAID defines a NPE as having significant barriers to operating effectively and safely due to one or more of the following factors:

- Armed conflict to which the U.S. is a party or not a party;
- Limited physical access due to distance, infrastructure, disaster, geography, or non-presence;
- Restricted political space due to repression of political activity and expression;
- Significant public health crises, such as a communicable disease outbreak or pandemic; or
- Uncontrolled criminality, including corruption.[14]

In such environments, BHA encourages partners to identify and pursue context-appropriate remote monitoring approaches that enable sufficient oversight and accountability of activity implementation,

[14] *ADS Chapter 201 Program Cycle Operational Policy*, USAID (2020).

including those discussed below. As remote monitoring is a rapidly evolving area of humanitarian M&E, BHA encourages open dialogue with partners to share best practices as they are developed in the field.

Adapting to Remote M&E Methods: Due to the nature of emergency activities, it is often necessary for partners to quickly adapt their monitoring approaches mid-implementation in response to heightened risks to staff and beneficiaries. For example, the global COVID-19 pandemic required many partners to rapidly adapt their monitoring and evaluation methods on a temporary, or in some cases a more prolonged basis, to comply with local public health ordinances and travel restrictions. In other contexts, the security situation may suddenly shift and preclude staff from performing in-person monitoring (e.g. routine monitoring or surveys) that were included in the Monitoring Approach.

BHA strongly recommends that partners plan ahead for possible contingencies and identify at the application stage any alternative monitoring methods, such as remote methods, that may be rapidly activated if needed. To this end, partners should consider which criteria they will use to determine when it is necessary to scale back in-person monitoring.

If the risk to staff or beneficiaries increases to a level that warrants a partner to adapt its planned monitoring approach mid-implementation, BHA recommends partners consider the following key principles:

- Prioritize "Do No Harm" for partner staff and beneficiaries.
- Pause or reduce monitoring of non-critical or non-life-saving activities, and revisit monitoring approaches regularly.
- Assess risk and burden on staff, communities, and beneficiaries of remote data collection.
- Update data collection tools and protocols to limit proximity, frequency and duration of face-to-face contact.
- Modify timeline or data collection methods for planned evaluations.
- Plan for capacity building and technical support for M&E staff and enumerators to ensure staff can execute modified and remote data collection methods.

When shifting to remote monitoring is not feasible, partners may use alternative methods to observe delivery of assistance or rely on observation methods that minimize direct contact (for instance, drive-by observations).

Partners should document and submit revisions to their award M&E Plans to reflect adapted M&E approaches for regular programming in response to any substantial changes in the operational context (e.g. outbreak of civil conflict or a global pandemic). These revisions should be submitted through the AOR who will circulate internally to the M&E team as appropriate for review and document the revision formally as part of the award documentation. Partners should also develop appropriate safety and supporting protocols that will be used for any remote or in-person data collection.

Finally, partners are also encouraged to plan ahead for a return to 'normal' operating conditions when it is safe and feasible to do so. For example, if a partner decided to switch to remote monitoring methods due to an exceptionally bad flood season in an area, they should already have plans in place for how to pivot back in in-person, direct data collection when the rainy season is over.

Third Party Monitoring (TPM) is one remote management tool that BHA and partners can use to monitor activities in NPEs. While the primary objective of TPM for emergency activities is to verify outputs, it can also be used to capture implementation challenges, successes, and community perception of the interventions. TPM involves contracting a third party organization to conduct both quantitative and qualitative data collection, through periodic site visits, remote (e.g. phone) and in person surveys, direct observation, and focus group discussions. Partners that elect to include TPM as part of their monitoring approach should describe the planned TPM methods in their M&E Plan at application, as well as any associated resources or budget allocated for management. TPMs should not replace a partner's internal monitoring systems, but as a complementary tool to assist in verification of activities in contexts where regular access may be limited.

In some high-risk contexts where partners use remote management and primarily operate through sub-awardees, BHA partners may ask partners to have their own TPM system as a risk-mitigation measure. This will be communicated to applicants during the application phase. BHA asks partners to follow these guiding principles for their TPMs:

1. Partners should prioritize third-party monitoring (TPM) site visits in areas where they do not have direct access or are implementing primarily through sub-partners.
2. The TPM must be conducted by a "third-party." They must be external to the partner or consortium.
3. The scope of the TPM should be limited, with a focus on output verification. Priority should be given to direct observations (e.g. distribution site visits) to observe whether activities are being implemented as planned and to receive feedback from beneficiaries.
4. The TPM contractor should report to the IP on a frequent enough basis to provide useful and timely information to project management. It is recommended that they report at least on a bi-monthly or monthly basis.
5. If a firm is selected for your TPM, share the name of the firm with your AOR. Coordination with your AOR will help to avoid potential conflicts of interest that may come if a partner sub-contracts to a firm that is also implementing BHA's TPM mechanism.
6. Partners should incorporate findings from the TPM into their regular reporting as outlined in the award.

Box 6. Cooperation with BHA-funded TPMs

Partners operating in countries where BHA utilizes a third-party mechanism will be expected to closely coordinate with the TPM contractor, and facilitate any requested site visits. Site visits typically involve the TPM contractor interviewing activity staff, key informants (community or camp leaders, etc.), and conducting focus group discussions with participants. Depending on the activity being monitored, TPM site visits will also include visual observations, such as adherence to warehousing standard operating procedures or observing the distribution process, or review of documents. Site visits are typically categorized using a rating system that indicates areas of concern, positive findings,

or the need for immediate action.[15] In addition to verifying outputs, TPM mechanisms may also monitor outcome indicators, such as FCS and coping strategies.

To facilitate TPM processes, partners are expected to provide timely responses to requests for information, including sharing activity documents, sharing/confirming current active site locations and intervention timing, and providing staff points of contact.

While the primary objective of BHA-funded TPMs is verification, it can also serve an important role as a complement to a partner's internal performance monitoring system. The results of each site visit are shared with partners for their awareness, and to respond to any issues that were flagged. This provides a useful flow of information about implementation that can help inform partner management decision-making.

Mobile Phone and Digital Data Collection: Partners may be able to plan ahead to set up a system for adapting in-person data collection instruments to phone-based interviews, web-based surveys, SMS, IVR.[16] Any introduction of alternative mobile or digital data collection technologies or platforms should ensure sufficient data security and privacy protocols are put in place by the partner to protect beneficiary PII or other sensitive data. Considerations for phone-based data collection:

- Shorten monitoring instruments to collect only essential information; reduce the number of questions being asked; reduce disaggregation requirements; and focus primarily on output-level indicator data.
- For interviews conducted by phone, consider using platforms/companies that do not charge the recipient for the airtime, and/or providing incentives in the form of cell phone credit, so beneficiaries do not have to use their own credit for the purpose of data collection. (Be sure to obtain and document verbal informed consent before beginning interviews.)
- For low-resource environments and those with limited cell phone penetration or ownership, consider identifying a trusted community liaison to equip with the appropriate technology to serve as an aggregator of data from the community.
- For partners already using mobile data collection systems, it is possible to extend services to embed direct messaging to clients (e.g., SMS surveys, Interactive Voice Recordings).[17]
- Partners should identify implications, risks and limitations of switching to phone-based data collection and identity mitigating measures, such as:
    - Fraud (e.g. the person on the phone is not the intended beneficiary);
    - Incomplete datasets as a result of call drops due to technical issues or respondents hanging up prematurely;
    - Response bias due to lower participation from vulnerable groups who may not have access to phones (e.g. women, girls, elderly, children, persons with disabilities);
    - Limited response due to lower cell phone penetration or service in certain areas;
    - Insufficient privacy for respondents answering questions in their home resulting in biased responses due to phone accessibility and "shared" or community phone;

[15] Note that the rating system will vary between TPM contracts, but should be well defined and communicated with all stakeholders.
[16] "Best Practices in Conducting Phone Surveys", J-PAL (2020).
[17] "CVA Payments and Digital Data Management- Deep Dive: COVID 19 and CVA", Mercy Corps (2020).

    - Potentially higher non-response rate via phone (consider refining sampling approach, such as over-sampling, to overcome this);
- Incomplete or unavailability of sampling frames for all sectors/sub-sectors, especially when IPs do not have telephone numbers from all the beneficiaries.
- Adjust training protocols for enumerators for phone data collection, including enhanced training to ensure informed consent, building rapport with respondents (especially for qualitative questions), and decreasing length of surveys with a focus on urgent questions.
- Partners should address and verify that sufficient levels of mobile connectivity and cell phone penetration exist in the operating area to ensure success of mobile methods; incentives or purchasing phone credits for respondents to complete phone surveys may also be considered and adequately budgeted.

Monitoring Through Key Informants[18]: In the case that beneficiaries cannot be reached by phone or mobile internet, monitoring through key informants (e.g., field-based project staff, extension workers, community health workers, non-governmental groups) may be an option if the key informants have access to SMS, voice calls, or mobile internet. If necessary, partner M&E specialists may be able to remotely train key informants to collect monitoring data. Most digital data collection apps are able to be used offline to collect data. This enables enumerators to collect information on their device while offline, and then send it at a later time, when the device has connectivity (e.g., on top of a hill, back at the regional office). Consider incorporating geolocated data to allow data quality checks.

Alternative Approaches to Beneficiary Verification: When it is not possible to track beneficiaries using signatures (e.g. for health reasons during a pandemic), partners may be able to use or switch to alternative technologies or other measures that may be effective to track participants without physical signatures, such as:

- Use GPS-enabled smartphones to take time-stamped and GPS-tagged photos of beneficiaries receiving the item during distribution, after receiving verbal consent from the beneficiary to have their photo taken.
- Conduct post-distribution monitoring by phone or video call to verify the items (e.g., food, NFI, hygiene kit) have been received by the intended beneficiary.
- Obtain informed consent verbally prior to collecting information by phone.
- Use Quick Response (QR) codes on the packaging of commodities, food and non-food items. Partner staff can use GPS-enabled mobile phones to scan the codes routinely throughout the delivery of the commodities to track their movement to the distribution endpoints.

Any technologies, digital platforms, or other methods employed should include **sufficient data security and privacy protocols**. Ideally these protocols should be put in place prior to implementing these practices, and verified on a regular basis to ensure PII and other sensitive data are protected.

For more resources on remote monitoring and adaptive management:

- Remote Food Security Monitoring Online Course: Introduction to Remote Data Collection Tools, WFP/mobile Vulnerability Analysis and Mapping (mVAM)(2017).
- Guide for Adopting Remote Monitoring Approaches During COVID-19, USAID/GDL (2020).

[18] Adapted from "USAID's Guide for Adopting Remote Monitoring Approaches During COVID-19", USAID (2020).

- Mobile Phone & Remote Tool Considerations for M&E in a COVID-19 Environment and Slides, Implementer-Led Design, Evidence, Analysis and Learning (IDEAL) activity, USAID (2020).
- Qualitative M&E During COVID-19: Sharing Tips for Remote Data Collection, FSN Network (2020).
- Best practices for conducting phone surveys, J-PAL (2020).
- Using mobile phone surveys to track resilience and post-disaster recovery: a how-to guide, ODI/BRACED (2020).
- Monitoring and accountability practices for remotely managed projects implemented in volatile operating environments, Tearfund (2012).

4.4 DATA QUALITY ASSURANCE

Ensuring the quality of data is a prime interest of BHA. Poor quality data can lead to wrong conclusions, undermine the need of the affected population, and performance of an activity. Poor quality data can misguide the implementing organization, BHA, the host country, and US tax payers. Given the difficult contexts and time-sensitive nature of emergency activities, partners must carefully design systems to ensure that data collected are of sufficiently high quality to meet management needs. The Monitoring Approach must describe how a partner will ensure that data collected and generated in their M&E systems meet the five key data quality attributes: **validity**, **reliability**, **timeliness**, **precision**, and **integrity**.

The Data Quality Assurance section of the Monitoring Approach should describe:

- Strategies used to reduce bias and errors in measurement, transcription, and processing of data. This should also include notes (either in each indicator PIRS and/or in the DQA section, as appropriate, on how double counting of individuals or households will be avoided).
- Documentation of methods and protocols for data collection, data entry and cleaning, coding, aggregation, and analysis.
- Procedures for verifying and validating the data collected by the M&E system. These procedures may include:
    - Site visits by project staff to participants who were respondents to surveys or another means of data collection in order to verify responses
    - Inclusion of photographs, video or audio recordings, or other evidence to allow others to verify observations, transcriptions, and interpretations by the collector[19]
    - Systematic review of collected data to compare values collected across time and location to flag outliers or reversals of trends that should be investigated
    - Incorporation of reasonability checks and comparisons into data collection, entry, and processing software; double keying of data in entry procedures; use of dropdowns and conditional entry fields; and developing filters, macros, and scripts to identify data outside reasonable parameters or data that contradict each other

[19] The DQA section should describe methods for safeguarding participant confidentiality when these methods are used.

Data Quality Assessments (DQAs) are periodic reviews to assess how effective the data quality assurance processes described in the monitoring plan have been at meeting the five key data quality attributes: validity, reliability, timeliness, precision, and integrity. The purpose of a DQA is to ensure that partners and BHA staff are aware of the strengths and weaknesses of indicator data, and the extent to which data integrity can be trusted to influence management decisions. A DQA is designed to:

- Verify the quality of the data
- Assess the system that produces the data
- Develop action plans to address identified issues and improve quality

DQAs can be particularly important for partners operating in non-permissive environments and implementing through remote management. The DQA can help the partners to identify threats to their data quality. BHA encourages all partners to complete one DQA during the course of the activity. For each DQA, BHA recommends that a partner focus on 2-3 key indicators. The selection of the indicators should be strategic, and may take into consideration:

- Indicators that are complicated to measure
- Indicators of suspect data quality
- Indicators of high importance to decision making
- Indicators that demonstrate an intervention's progress
- Indicators that represent different data flow processes

In the DQA, reviewers will review the flow of data for each of the selected indicators to verify their quality and potential sources of error at each stage, beginning from the initial point of collection and continuing through reporting and use. The DQA process may examine:

- M&E staffing, functions, and capabilities
- Indicator definitions and reporting guidelines
- Data collection tools and reporting forms
- Processes of data verification, aggregation, processing, management, storage, and safeguarding
- Data use and dissemination practices

For partners planning to conduct a DQA, the Monitoring Approach should describe the timing and processes, including:

- A list of indicators to be reviewed and a justification for the selection
- Timing and duration of the planned DQA
- Specific focus of the review (e.g. identify a particular step in the data collection process that has been identified as a risk to data quality)
- Roles and responsibilities for conducting the DQA

A DQA will typically be implemented by the partner (internal DQA). When an internal DQA is conducted, it should be led by someone who is not directly responsible for collecting the data that is being assessed, such as a regional or head-office M&E advisor. DQAs can also be externally conducted for increased independence. The findings from any DQA should be shared with the activity's management, and should be used to improve the data collection processes and systems for the selected

indicators. BHA strongly encourages partners to also share DQA findings with BHA. USAID may choose to conduct its own DQA, which may be conducted by BHA M&E staff or by a contractor.

For more information on DQAs, see the following resources:

- ADS 201.3.5.8, USAID (2020).
- BHA DQA Webinar handout, USAID/FANTA/FHI360 (2016).
- MEASURE Evaluation Data Quality Assessment Methodology and Tools, MEASURE Evaluation

Box 7. Primary and Secondary Data Quality

While collecting primary data requires more time and resources, partners have significantly more control over the quality of primary data. Secondary data are data collected by someone else for a different purpose, so partners should be sure to check the quality of secondary data before using it for monitoring or evaluation.

4.5 DATA MANAGEMENT AND SAFEGUARDING

The Monitoring Approach must describe a partner's plan for protecting data from unintended change, misuse, loss, or destruction as it is collected and as it flows between and through the various sites of processing to its final storage location. This relates to data on paper, on other media, and in digital format. Any breach of privacy or inappropriate use of data can potentially result in negative unintended consequences, especially in contexts with conflict or internal divisions and tensions. Therefore, access to data for viewing, use, and modification must be restricted. The plan should also describe how and for how long the data will be preserved for future use. For consortium or partnership activities, the Monitoring Plan must describe how data management will be coordinated across partners.

Examples of data management and safeguards include:

- Measures that will be taken to ensure and safeguard participant confidentiality and protect personal identity information, including on both hard copy and digital files
- Systems to store/maintain original data files/activity records: Where original data will be stored, how they will be protected, who can access them, how long the partner will retain them, and procedures and timeline for their destruction
- Methods, frequency, and locations of file and database backups and who is responsible for making backups; measures to prevent and detect unauthorized data access for data entry, editing, processing, or retrieval; virus protection of digital data; and security measures to protect the physical location of hard copies, databases, and data backups

CHAPTER 5: BASELINE AND ENDLINE

5.1 BASELINE/ENDLINE STUDY REQUIREMENTS

A baseline study is required for all awards that are six months or longer in duration, and must be submitted to BHA within 90 days of award approval. For longer awards or those using more complex baseline methodologies, partners may submit written justification to the AOR to request an extension on the baseline report deadline. The baseline may be conducted by the partner directly if qualified staff are available, or contracted to a qualified third-party firm to implement the study.

The purpose of the baseline study is to collect data for all indicators included in the ITT before implementation begins. Baselines should also collect non-indicator information to describe the prevailing conditions of the target communities or population. Baseline values serve as a point of comparison with endline values during the final evaluation. They also provide the partner with important information about their affected population that can be used to improve targeting and activity design before implementation begins. In many cases, the baseline study will represent the most thorough recent study of the target population and can provide valuable insights to activity staff.

The baseline study must collect data on:

- All BHA and custom outcome indicators included in the applicant's ITT. They must be collected and calculated exactly as described in the PIRS. These indicators should not be modified or substituted without approval from BHA (AOR and M&E Advisor).
- Non-indicator information to describe the prevailing conditions of the target communities or population, including community and/or household characteristics. Include key findings by sector and sub-sector, including location-specific assessments for shelter, protection or health facilities that could not be conducted in advance of the activity. Baseline reports may build on previous needs assessments, but include more specific information on the target communities the partner will be working with.
- For output indicators with a baseline value of 0, partners may reference the monitoring approaches they will use to collect data for that indicator throughout the life of the activity.

Some activities report primarily output indicators with zero-value baselines. In cases where the baseline value of many indicators is zero, baseline data collection and analysis will be less complex (e.g., they will likely not report on outcome indicators that require probabilistic sampling techniques or surveys) and rely on more rapid qualitative data collection and updated needs assessments to describe the prevailing conditions of the target beneficiaries and locations.

Box 8. Summary of Baseline Study Requirements

When a Baseline Study is Required

- Required for awards six months or longer
- Optional for awards shorter than six months

Who Conducts Baseline

- Partner or external firm

Requirements for Partners Conducting a Baseline

- Submit abbreviated statement of work (SOW) with the application
- Complete data collection **within 90 days** of award approval
- Submit final Baseline report in BHA ART for AOR approval and update baseline targets and values **within 90 days** of award approval
- Submit final Baseline report to the DEC[20]
- Submit all datasets to the DDL in accordance with ADS 579

Endline data are the final life of award (LOA) values for all activity indicators collected at the end of an activity. Endline data collection serves one primary purpose: to provide a comparison to the baseline value. You must submit endline data for all activity indicators as part of the final performance report, uploaded into BHA ART within 90 days of the end of the award. Regardless of whether an evaluation is planned, endline values for all indicators must be collected and reported at the end of the activity. If you plan to do a final evaluation, you may include endline data in the final evaluation report, in addition to the final report.

BHA distinction between endlines and evaluation: BHA accepts evaluations that are conducted during the course of implementation (mid-term evaluations, real-time evaluations) or at the end of an activity (final evaluations). Often a final evaluation may include endline data collection, but not exclusively. Evaluations seek to answer a breadth of questions, which go beyond only measuring the final indicator values at the end of an activity. It is often guided by evaluation questions oriented around OECD-DAC criteria. Evaluations may be qualitative in cases where statistically comparable baseline/endline surveys are not appropriate, necessary or feasible based on the indicators for that activity. BHA expects partners to propose an evaluation design that is appropriate to their proposed intervention.

Baseline Requirements for Follow-on Awards

If a partner has back-to-back awards implementing the same interventions, among the same cohorts of beneficiaries in the same geographic areas and is reporting on the same indicators, BHA encourages partners to consider whether it is appropriate to use the endline values from relevant indicators generated the previous award as the baseline for their new follow-on award.

For outcome indicators the endline value of the previous award may be used as the baseline for the follow on award only if the following conditions are met:

a) The intervention is targeting the same geographic locations and population

[20] In a few select countries that BHA supports there is a waiver that removes requirements for submission of reports to the DEC; the AOR for each award will inform partners if a DEC waiver is in place.

b) The partner is able to employ the same sampling frame and methodology for endline data collection, to allow for comparison between the previous award and follow-on.

If both of these conditions are not met, then new baseline data should be collected at the start of the follow on award for any new target populations and geographic locations.

For output indicators, baseline values for follow on awards should be zero.

5.2 BASELINE/ENDLINE DATA COLLECTION METHODS

Baseline/endline data collection may employ a variety of quantitative and qualitative methods. Methods should be appropriate, cost efficient,and in line with humanitarian principles. BHA generally prefers primary data collection, although secondary data are permissible where operational context may not allow for primary data collection. Data collection methods should be determined by the indicators the partner is collecting**,** adhering to what the BHA PIRS prescribes for each indicator. Refer to the Data Collection section of each PIRS in the BHA PIRS Handbook, as well as Chapter 3 for guidance on selecting a Data Collection Method.

Quantitative methods: For indicators that are measured through beneficiary-based or population-based surveys, BHA requires a probability sample (see note below for exceptions). The sample sizes for the surveys should be designed to detect statistically significant changes in estimates from baseline to endline. See Chapter 3 for guidance on Sampling for probability-based surveys.

Note: Exceptions to Baseline/Endline Representative Household Surveys

Activity less than 12 months: In general, BHA does not encourage partners to conduct representative surveys at baseline/endline for activities less than 12 months in duration. If an activity is required to report on a BHA outcome indicator per the PIRS that is measured through representative household survey, but the partner does not anticipate affecting this level of change in a shorter-term intervention or the partner believes that conducting the survey will be overly burdensome, the partner may provide justification in its application M&E Plan for BHA review to either 1) omit this indicator from the M&E Plan, or b) replace baseline/endline survey with enhanced PDM that includes outcome monitoring.

Replacing Baseline/Endline Survey with Enhanced PDM: For activities less than 12 months or in cases where an activity works with multiple cohorts of participants with short-term interventions (e.g. 1-3 months of rations or cash transfers), the requirement for measuring outcome indicators (e.g. FCS, rCSI, HHS) at baseline/endline via representative surveys may be waived in lieu of a more robust PDM survey that includes outcome monitoring as part of registration and PDM shortly after the final transfer. This option allows the partner to measure outcome indicators as part of implementation without launching a separate baseline/endline survey exercise.

Qualitative methods: Emergency and DRR activities may employ non-survey methods as part of baseline data collection to measure indicators and/or collect information on the prevailing conditions of the target population. Illustrative methods include:

- Systematic assessment of targeted health facilities.
- Document review of health facility registers to ascertain prevalent health issues in the target area
- Pre-tests to measure individual knowledge acquisition before training for DRR, HCIM, or other sector-specific training
- Technical assessments of damaged shelters to be rehabilitated
- Desk review of existing policies, early warning systems, and procedures that the partner seeks to strengthen through the planned DRR intervention
- Organizational capacity assessment with DRR stakeholders or local NGOs to inform capacity development plan
- Water quality testing at communal water points to be rehabilitated through the activity

5.5 BASELINE/ENDLINE STUDY TIMING

Data collection for the baseline study must be completed and submitted to the AOR within 90 days of the approval of the award. Following AOR approval, the baseline study should be submitted to the DEC. Data collection should take place before implementation has begun in order to get an accurate measure of participants' baseline status. However, the emergency context and timing may require data collection to take place concurrent with the start of the intervention.

It is important to ensure that data is collected quickly so as to not delay implementation. However, implementation should not wait if the partner cannot conduct the baseline right away. If implementation begins before the baseline data is collected, this should be discussed in the "limitations" section of the baseline study report. Ideally, baseline and endline data should be collected during the same season to ensure comparability of data - particularly for food security and agriculture programs. BHA recognizes that this may not be possible for shorter awards or other challenges. Partners are encouraged to address any implications of not matching seasonality in the limitations section of their baseline report, as well as in the final performance report (and evaluation report, if applicable).

For best results, the endline survey should be conducted directly after the intervention has ended. To ensure comparability of data between baseline and endline, the endline should ideally be conducted in the same season as the baseline, though this may not be possible for awards of certain lengths (e.g. 18 months). In cases where it is not possible to collect data directly following the end of the intervention AND during the same season as that for the baseline data collection priority should be given to completing the endline data collection as close to the end of the intervention as possible.

5.5.1 BASELINE INTEGRATED WITH ROUTINE MONITORING

For collecting baseline data on indicators that require a quantitative survey, BHA encourages partners to consider using the beneficiary registration process as a means of baseline data collection. If all required indicators can be collected from either all households or a representative subset of households during registration, it can save time and resources that would otherwise be devoted to a separate survey. For instance, for a multipurpose cash assistance (MPCA) or food distribution activity, baseline values for

outcome indicators may be collected as part of the enrollment or registration process. In this case endline data may be collected during the final PDM so long as the sampling methodology for PDM surveys that include outcome indicators enable statistical comparison (95% confidence interval, 5% margin of error) between baseline and endline.

5.5.2 BASELINE FOR ACTIVITIES WITH ROLLING ENROLLMENT ("ROLLING BASELINES")

Many emergency activities enroll new participants on a rolling basis throughout implementation rather than all at once at the start of the activity. This is common for interventions that involve distributions, such as food assistance, MPCA, or NFIs, or activities involving multiple cohorts for training/capacity building. In these cases, it is common for partners to capture baseline characteristics of each cohort as they are enrolled, either through a beneficiary-based survey or a beneficiary census. This can complicate the collection of baseline data because not all beneficiaries may have been identified in the first 90 days of the activity, and has implications on the sampling approach and timing of baseline and endline surveys.

Figure 5 below shows three different scenarios of rolling enrollment that have implications for data collection. These examples are not comprehensive, but illustrate key considerations.

Figure 5: Scenarios to Consider with Rolling Enrollment

Scenario 1: No overlap	Q1	Q2	Q3	Q4
Cohort 1				
Cohort 2				
Cohort 3				
Cohort 4				

Scenario 2: Overlap	Q1	Q2	Q3	Q4
Cohort 1				
Cohort 2				
Cohort 3				
Cohort 4				

Scenario 3: Phased inclusion	Q1	Q2	Q3	Q4
Cohort 1				
Cohort 2				
Cohort 3				
Cohort 4				

In **scenario 1**, there is no overlap between cohorts. Each receives short-term assistance (for instance, 1-3 months of food rations, one months' worth of hygiene kits, one-time MPCA transfer). In **scenario 2**, each cohort receives short-term assistance, with some overlap between cohorts. Note: In cases where an activity works with multiple cohorts of participants with short-term interventions (e.g. 1-3 months of rations), the requirement for measuring outcome indicators (e.g. FCS, rCSI, HHS) at baseline/endline via representative surveys may be waived in lieu of a more robust PDM that includes outcome monitoring. For those that do conduct baselines and endlines, the timing of the data collection should be considered. Baseline data is typically collected as beneficiaries are enrolled. Endline data should typically be collected at a consistent interval of time after the final distribution. If instead baseline data is collected on a sample of all participants at the end of the activity (e.g., end of Q4), then some participants may still be receiving assistance (e.g., cohort 4 in scenario 1), while others will have gone many months without assistance (e.g., cohort 1 in scenario 1). This can complicate the interpretation of the findings.

In **scenario 3**, households are enrolled on a rolling basis, but receive continuous assistance through the life of the activity once enrolled. In this scenario, it is likely that the baseline will be taken at enrollment and the endline will be taken on a sample of all participants at the end of the activity. The analysis and interpretation of findings will need to take into account that the baseline data was collected across different seasons, and that participants received assistance for different lengths of time before endline.

Given the challenges with conducting baseline and endline surveys for activities with rolling enrollment, partners implementing activities that work with multiple cohorts of participants with short-term interventions (e.g., scenarios 1 and 2) are encouraged to propose other approaches to measure outcomes, such as a round of PDM that includes outcome indicators after the final distribution.

The methods used for rolling baselines should be clearly communicated in the Abbreviated Statement of Work for Baseline/Endline, including assumptions made when aggregating data from different cohorts of participants. If participants are enrolled at different times of the year, there could be differences in baseline characteristics due to seasonality, but sample sizes will likely not be large enough to test for differences between cohorts.

If the activity will use rolling beneficiary registration or cohorts, baseline data collection should be collected on a rolling basis or for each cohort. Partners conducting rolling baselines that will continue beyond the first 90 days of implementation should discuss with the AOR and BHA M&E advisor an appropriate timeline for submission of the baseline report, including DEC and DDL submissions. BHA may request that the partner submit an initial round of baseline data within the first 90 days of the award, per the award requirement, explaining in the narrative report its plan for how baseline data for later cohorts will be collected. Baseline data and analysis for subsequent cohorts may be submitted as part of subsequent semi-annual reports, or as otherwise agreed with the AOR.

5.6 BASELINE STUDY REPORT

Partners implementing activities with a duration of six months and longer must submit a baseline report, including an updated ITT with actual baseline values and updated targets and PIRS for custom indicators, into BHA ART within 90 days of the award. For longer awards or those using more complex baseline

methodologies, partners may submit written justification to the AOR to request an extension on the baseline report deadline.

BHA encourages partners to be as concise as possible (maximum length 10 pages, excluding annexes). The baseline report should be appropriate to the scope and complexity of the award. BHA provides a suggested report outline in Annex 4. Baseline Report Suggested Format.

5.7 USE OF BASELINE STUDY RESULTS TO REFINE ACTIVITY STRATEGIES AND INDICATOR TARGETS

Baseline studies often represent the most in-depth and recent study of the target population. BHA expects that partners will use the baseline study results to review their activity design and refine implementation as necessary. For example, an activity with an IYCF component may find that the prevalence of children 6-23 months receiving a minimum acceptable diet was much lower than anticipated, and decide to re-allocate more resources to their activities working to improve this. Partners can consider holding a workshop to present their baseline findings to staff, discuss assumptions that may have been challenged, and identify how implementation should be adjusted.

Baseline findings may reveal the need to update performance indicator targets that were included in the application. If a partner proposes to revise one or more performance indicator targets based on baseline findings, they should seek AOR concurrence through the following process:

- Update the "Target" column of the ITT and submit as an annex to the baseline report
- Provide justification for each indicator target revisions in the narrative baseline report
- If the AOR concurs with the revised targets, the partner should update BHA ART with the revised indicator targets.

Note: A formal award modification is not needed to update indicator targets in the ITT. However, updates to the total number of target beneficiaries may require an award modification and should be discussed with the AOR separately.

CHAPTER 6. EVALUATION

Evaluation plays an important role in fulfilling BHA's obligation to ensure the effective and efficient use of resources as a tool for both accountability and learning. This chapter provides guidance on BHA's requirements for evaluations for emergency activities.

6.1 WHEN TO EVALUATE

One of the most important considerations when planning an evaluation is deciding which activities to evaluate and when to conduct the evaluation. BHA encourages applications to plan strategically for evaluations that will provide useful evidence to inform decision-making.

BHA requires evaluations under the circumstances described in Box 9. below. Applicants are encouraged to propose evaluations when not required by BHA. An Evaluation Approach must be submitted as part of the M&E Plan at application when an evaluation is planned.

Box 9. Summary of Evaluation Requirements

When an Evaluation is Required

- If the original period of performance for the activity is 18 months or longer
- If your organization has implemented at least one BHA-funded award (of any duration, in any sector) in the past three years in a given country and your organization has not completed an evaluation of any BHA-funded awards in that given country in the past three years. Partners must complete at least one evaluation of any BHA-funded award(s) at least once every three years in a given country.

Who Conducts the Evaluation

- Either a third-party firm, or an internal team led by an experienced external team leader

Requirements for Partners Conducting an Evaluation

- Submit abbreviated statement of work (SoW) with the application
- Submit a full SOW six months prior to the start of the evaluation
- Submit final evaluation report to AOR and DEC
- Submit all datasets to the DDL in accordance with ADS 579

6.2 EVALUATION PURPOSE & QUESTIONS

6.2.1 EVALUATION PURPOSE

There are two primary purposes of evaluations: learning and accountability. In reality, most evaluations will serve a dual purpose. It is important to consider the purpose and audience of the evaluation as a first step in the evaluation planning process. The evaluation questions, methods, and timing should be carefully selected to fulfill the specific purpose of the evaluation.

6.2.2 EVALUATION QUESTIONS

When drafting the evaluation SOW, it is important to ensure that evaluation questions are consistent with the evaluation objectives, and that the evaluation methods are appropriate for answering the evaluation questions. It is also important to structure the evaluation to the context of the activity. For example, an evaluation of a response to a sudden-onset emergency (e.g., earthquake or flood) should look different to a response to a protracted crisis (e.g., protracted IDP crisis). For example, a partner implementing a shorter activity responding to a sudden-onset emergency may choose to conduct a simple qualitative evaluation focused primarily on operational lessons-learned, while a partner implementing an activity in a protracted crisis may conduct a mixed-methods evaluation utilizing baseline and endline household survey data to measure changes in outcomes.

BHA typically recommends that between one and five evaluation questions are selected, and that each evaluation question is concise with well-defined terms. Avoid long lists of poorly-defined or difficult-to-answer questions. Keep in mind that the evaluation questions should focus on what is most important - not every aspect of an activity needs to be evaluated. Vague terms like "relevance" and "effectiveness" can be interpreted in many ways, so clear definitions should be provided. Evaluation questions should be listed by order of importance, with the first question being the most important.

The following list of illustrative evaluation questions, organized by the OECD/DAC Evaluation Criteria, can be used as a reference when drafting the SOW.[21] BHA does not expect each evaluation to address all criteria; the partner should select questions that are most relevant to their learning needs.

a) **Relevance: Is the intervention doing the right things?**
    - Were interventions appropriate and effective for the target group based on their needs?
    - Which target groups and individuals were reached by the interventions?
    - How effective was the targeting approach in achieving the activity goal?
b) **Coherence: How well does the intervention fit?**
    - To what extent did the activity consider gender equity, protection, age, physical and emotional challenges of the participants, and risks to participation in various interventions in activity design and implementation?
    - How has management adapted the activity design or implementation based on monitoring information and feedback from the target population?
c) **Effectiveness: Is the intervention achieving its objectives?**[22]
    - To what extent do the activity's interventions appear to have achieved their intended outputs and outcomes?
    - To what extent did the activity help prevent individuals and households from adopting negative coping strategies such as selling productive assets?
d) **Efficiency: How well are resources being used?**
    - How were problems and challenges managed?

[21] "OECD Evaluation Criteria", OECD (2020).

[22] Performance evaluations do not contain a rigorously defined counterfactual, so they should not answer questions about the amount of change in outcomes directly attributable to an intervention.

- To what extent have the activity's interventions adhered to planned implementation schedules?
- What was the level of efficiency and timely delivery of the goods or services?

e) Impact: What difference does the intervention make?

- What changes—expected and unexpected, positive and negative—were experienced by the targeted beneficiaries and other stakeholders?
- What factors appear to facilitate or inhibit these changes?
- Which interventions appeared to be more or less important to achieving activity outcomes?
- How did these changes correspond to those hypothesized by the activity's Theory of Change?

f) Sustainability: Will the benefits last?

- To what extent did the activity take advantage of other USG and non-USG investments in the same target areas to facilitate linkages with complementary services, layering with earlier investments, and implementing an exit strategy?
- To what extent did the activity align and integrate with host government social protection strategy/policy/service delivery?
- Was the activity able to end operations at the close of the award without causing significant disruptions in the targeted communities?

6.3 EVALUATION TYPES & METHODS

BHA supports a range of evaluation types. The type of evaluation selected must be appropriate to answer your evaluation questions. Evaluations fit broadly into two categories: performance evaluations, and impact evaluation. The following definitions come from USAID's Evaluation Policy.[23]

> ***Performance evaluations*** *encompass a broad range of evaluation methods. They often incorporate before-after comparisons, but generally lack a rigorously defined counterfactual. Performance evaluations may address descriptive, normative, and/or cause-and-effect questions. As performance evaluations do not contain a rigorously defined counterfactual, they should not answer questions about the amount of change attributable to an intervention, where other factors are likely to have influenced the variable in question.*
>
> ***Impact evaluations*** *measure the change in an outcome or a set of outcomes that is attributable to a defined intervention. Impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a treatment or a control group provide the strongest evidence of a relationship between the intervention under study and the outcome measured. Impact evaluations may use an experimental or a quasi-experimental design.*

The majority of evaluations conducted for BHA-funded emergency activities will fall under the performance evaluation category. Examples of evaluation types are described below.

[23] *USAID Evaluation Policy,* USAID (2016).

6.3.1 PERFORMANCE EVALUATIONS

Mixed-methods Performance Evaluations consist of both quantitative and qualitative data collection, which are systematically integrated. A final mixed-methods performance evaluation should integrate a comparison of baseline and endline quantitative data, as well as a qualitative study. The qualitative study should be designed to explore issues identified in the quantitative results and answer evaluation questions that are beyond the scope of the quantitative survey (e.g., sustainability, management, etc.). Where possible, mixed-methods performance evaluation should pull from other sources of data including the activity's performance monitoring data.

Examples of mixed-methods performance evaluations include:

- A **mixed-method midterm evaluation** may look at process-evaluation questions related to the quality of implementation, while incorporating quantitative survey data.
- A **mixed-method final evaluation** will integrate a comparison of baseline and endline quantitative data, as well as a qualitative study. The performance evaluation may also include a review of performance monitoring data.

Qualitative Performance Evaluations use a range of qualitative methods to answer evaluation questions which should be selected in order to accurately answer the evaluation questions. These methods and protocols should be designed to ensure that if a different, well-qualified evaluator were to undertake the same evaluation, he or she would arrive at the same or similar conclusions. A variety of primary data collection methods should be used, including: semi-structured and in-depth interviews, focus group discussions, and direct observations.

Examples of qualitative performance evaluations include:

- A **qualitative midterm evaluation** will objectively review the progress of implementation, assess implementation quality, identify challenges faced, and provide recommendations for course correction.
- A **qualitative final evaluation** will objectively review the activity's achievements against plans, assess implementation quality, identify challenges faced, and provide recommendations for future activities.

Note that **Real-Time Evaluations (RTEs)** are also typically conducted using qualitative methods, and may be supported by BHA under certain circumstances though they may not fulfill BHA's evaluation requirement if they do not specifically evaluate the BHA-funded activity. RTEs are evaluations of an ongoing humanitarian response, typically conducted early-on in the response (typically within the first three months). RTEs typically rely on qualitative methods, and are designed to provide rapid feedback in order to improve operations or course-correct.

6.3.2 IMPACT EVALUATIONS

BHA may support an impact evaluation, especially when the applicant provides a sufficient justification for the impact evaluation filling a critical evidence gap. The applicant must also document that they have sufficiently considered and addressed the logistical and ethical considerations that come with conducting an impact evaluation in a humanitarian context. The objective of an impact evaluation of a humanitarian

assistance activity should be to fill gaps in evidence that will lead to more effective and efficient humanitarian responses. Where possible, the evaluations should attempt to answer practical implementation questions about comparative cost-efficiency of different interventions or approaches. The evaluations may use both experimental or quasi-experimental design. The methods should be appropriate to answer the evaluation questions given the operating context.

Experimental Impact Evaluations use random assignment to select treatment and control groups from the targeted population. These evaluations are often referred to as randomized-controlled trials (RCTs) due to the process for assigning treatment and control groups. Experimental impact evaluations provide the strongest evidence of impact, and are especially effective at addressing issues of selection bias. Because of this, the results are often simpler to analyze and interpret than for quasi-experimental impact evaluations. At the same time, they can be challenging to implement, especially in humanitarian contexts. There are a number of experimental impact evaluation approaches, including simple random assignment, randomized phase-in, and multiple treatments.

Quasi-experimental Impact Evaluations use statistical methods to estimate the counterfactual where random assignment is not possible. Common quasi-experimental methods include matching and regression discontinuity. There are a number of different matching approaches, with propensity-score matching (PSM) being among the most commonly used. Matching approaches rely on selecting comparison groups by matching on observable characteristics. Regression discontinuity design (RDD) is an approach that is appropriate for activities that have clear targeting criteria with a cut-off that determines who is eligible to participate. Outcomes of beneficiaries and non-beneifices just above and below the cut-off are compared.

6.3.3 QUANTITATIVE EVALUATION METHODS

Quantitative data for most evaluations should come from the baseline and endline data collection following the methods described in Chapter 5. The quantitative methods used must be consistent with the requirements described in the PIRS for the indicators that will be measured, and must be appropriate for the evaluation type. Partners should closely coordinate their baseline/endline data collection with the evaluation team where a mixed-method performance evaluation is planned. In some cases, baseline and/or endline data may be collected by an external firm. When an impact evaluation is planned, the evaluators must be consulted as soon as possible to collaborate with the partner on the baseline design and data collection.

6.3.4 QUALITATIVE EVALUATION METHODS

Evaluations may utilize a range of qualitative methods including semi-structured interviews, in-depth interviews, focus group discussions, and direct observations. There should be a clear plan for sampling to ensure that a range of different stakeholders are consulted from different geographic areas.

6.4 EVALUATION SOW

Applicants planning to conduct an evaluation must submit an Evaluation Approach component of their M&E Plan, which includes an abbreviated Statement of Work (SOW). This abbreviated SOW should succinctly document the following:

- Evaluation purpose
- Evaluation type
- Evaluation questions
- Evaluation methods
- Evaluation timeline
- Dissemination plan
- Evaluator profile

Detailed guidance on the requirement can be found in *Annex 3: Guidance for Abbreviated Statement of Work for Evaluations*. The abbreviated SOW submitted at application will allow BHA to assess the appropriateness of the proposed evaluation. While this should represent the best estimate of what will be evaluated at the time of application, these plans may evolve as implementation begins. BHA requires a full evaluation SOW at least six months prior to the start of the evaluation.

6.5 EVALUATION REPORT

The evaluation team leader is responsible for drafting the final evaluation report. It is important to ensure that both the quantitative and qualitative components are well-integrated and are used to support cohesive findings. BHA expects that evaluation reports will be well-written, insightful, and concise. Once the report is finalized it should be submitted to the AOR along with the final project report, and then uploaded to the DEC.

All evaluation reports should be formatted consistently with USAID's evaluation report template.

Resources:

- Evaluation of Humanitarian Action, ALNAP (2016).
- Real-Time Evaluations of Humanitarian Action, ALNAP (2009).
- Evaluation Criteria, OECD/DAC
- Technical Note: Conducting Mixed-Methods Evaluations, USAID/PPL (2013).
- Technical Note: Impact Evaluations, USAID/PPL (2019).

CHAPTER 7: REPORTING & CLOSEOUT

This chapter provides a summary of M&E reporting and closeout requirements for BHA emergency awards. This should serve as a supplementary resource. Partners should reference the language included in their award document as their primary source of information. In addition, partners should refer to the award document for additional BHA financial reporting and activity closeout.

7.1 REPORTING

In any given FY, the partner will only submit at most two programmatic performance reports. There are three types of programmatic performance reports: semi-annual report, annual report and final performance report. For every semi-annual reporting period, the partner will provide semi-annual and unique FY values. At the end of the award, in addition, the partner will provide LOA values.

The purpose of the programmatic performance reports are to share progress against indicators identified in the Recipient's M&E Plan. The programmatic reports must tell the story behind the indicator, and share any planned changes in programmatic approaches. As applicable, BHA requires post distribution monitoring (PDM) narrative related to distributions and transfers (i.e., food, non- food items, in-kind, cash, and vouchers), and the role of the goods in achieving the activity purpose(s) and outcomes. The PDM narrative should describe satisfaction with the process of distributions and with the transfers received, as well as beneficiary perspectives on the outcomes of the distributions and transfers. See other specific requirements based on the modality may be included in the award documentation. See also Annex 5 for additional information on a suggested format for reporting.

7.1.1 SEMI-ANNUAL PERFORMANCE REPORT

All emergency activities must submit semi-annual reports (SAR) within 30 days after the end of FY Q2 (no later than April 30) regardless of when the award was awarded, except no SAR is required when the award start date is within 60 calendar days prior to the SAR due date of April 30. In those cases, the period(s) constituting the exception must be covered in the following semi-annual reporting period (Annual Report). As appropriate, update all activity baseline indicators in BHA ART and the ITT. SAR narrative and all annexes must be uploaded and all required and required if applicable indicator values via BHA ART. See requirements summarized in the table below and Annex 5 for additional information on a suggested format for reporting.

7.1.2 ANNUAL PERFORMANCE REPORT

The annual report (AR) is due within 30 days after the end of FY Q4 (no later than October 30) regardless of when the award was awarded. The AR is a fiscal year annual reporting requirement for all current BHA awards implemented by U.S. or non-U.S. non-governmental organizations (NGOs), including private voluntary organizations (PVOs). Partners must submit an annual narrative report covering the performance period October 1 - September 30, provide direct data entry of semi-annual and unique FY values and upload other documents as outlined in table below. See also Annex 5 for additional information on a suggested format for reporting.

The partner is not required to submit both an Annual Report and a Final Performance Report for the same reporting period in the final fiscal year of an award. In the case the award end date is in FY Q4, the Recipient must submit the Final Performance Report in lieu of the AR on the AR due date (no later than October 30).

7.1.3 FINAL PERFORMANCE REPORT

Final performance reports (FPR) are due 90 calendar days after the award end date and the requirements are summarized in the table below, articulated in the award language and described in the Suggested Format for Reporting in Annex 5. Recipients must submit a life of award narrative report covering the performance period of the award, provide direct data entry of semi-annual and unique FY and LOA values and upload other documents as outlined in the table below.

Table 7. BHA Emergency Narrative Report Suggested Format

SEMI-ANNUAL PERFORMANCE REPORT	ANNUAL PERFORMANCE REPORT	FINAL PERFORMANCE REPORT
1. Overall Performance 2. Changes and Amendments 3. Measuring Results (see below for additional requirements) 4. Participation & Accountability to Affected Populations (AAP) 5. Risk Management 6. Coordination 7. Challenges and Proposed Solutions 8. Planned Interventions	1. Overall Performance 2. Changes and Amendments 3. Measuring Results (see below for additional requirements) 4. Participation & AAP 5. Risk Management 6. Coordination 7. Challenges and Proposed Solutions 8. Planned Interventions (if applicable) 9. Exit Strategy & Sustainability (if applicable)	1. Overall Performance 2. Changes and Amendments 3. Measuring Results (see below for additional requirements) 4. Participation & AAP 5. Risk Management 6. Coordination 7. Challenges and Proposed Solutions 8. Exit Strategy & Sustainability

Table 8. Additional BHA Emergency Reporting Components

SEMI-ANNUAL PERFORMANCE REPORT	ANNUAL PERFORMANCE REPORT	FINAL PERFORMANCE REPORT
UPLOAD DOCUMENTS 1. Narrative Report (Required) 2. Indicator Tracking Table (Required) 3. Baseline, Evaluation, Assessments, and Research Reports (RiA) 4. Success Stories (recommended) **DIRECT DATA ENTRY (Semi Annual Values)** 5. Activity level Unique Beneficiaries 6. Sector level Unique Beneficiaries 7. Activity level Unique Refugee and IDP Beneficiaries 8. Emergency indicators 9. Baseline values	**UPLOAD DOCUMENTS** 1. FY Narrative Report (Required) 2. Indicator Tracking Table (Required) 3. Baseline, Evaluation, Assessments, and Research Reports (RiA) 4. Success Stories (recommended) **DIRECT DATA ENTRY (Semi Annual & FY Values)** 5. Activity level Unique Beneficiaries 6. Sector level Unique Beneficiaries 7. Activity level Unique Refugee and IDP Beneficiaries 8. Emergency indicators	**UPLOAD DOCUMENTS** 1. LOA Narrative Report (Required) 2. Indicator Tracking Table (Required) 3. Baseline, Evaluation, Assessments, and Research Reports (RiA) 4. Success Stories (recommended) **DIRECT DATA ENTRY (Semi Annual, FY & LOA Values)** 5. Activity level Unique Beneficiaries (semi-annual data) 6. Sector level Unique Beneficiaries (semi-annual data) 7. Activity level Unique Refugee and IDP Beneficiaries (semi-annual data) 8. Emergency indicators **(Required if Applicable)** 9. Endline/Evaluation values 10. LOA Actuals Data Tables • LRIP Procurement • Modality Actuals

Note: All values are unique counts that avoid double counting.

7.2 CLOSEOUT

7.2.1. SUBMISSION OF REPORTS TO THE DEC

The Development Experience Clearinghouse (DEC) is the largest online resource for USAID-funded technical and activity materials. PVO partners are required to submit documentation created during the course of their award to the DEC, such as assessments, analyses, studies, articles, baselines studies, midterm and final evaluation reports. The partner must provide the AOR with a DEC link or screen shot of the submitted document as proof of submission Partners should review their award language

and consult with their AOR if there are questions about what must be submitted to the DEC. If necessary, seek approval from AOR for a DEC waiver.

Resources:

- USAID's ADS 540: USAID Development Experience provides policy directives, required procedures, and roles and responsibilities governing the submission of materials to the DEC.

7.2.2. SUBMISSION OF DATA TO THE DDL

The Development Data Library (DDL) is the Agency's repository of USAID-funded, machine readable and non-proprietary data created or collected by the Agency and its partners. According to ADS 579, any dataset created or collected with USAID funding must be submitted to the DDL. This includes datasets produced by the partner and its sub-partners/contractors. For BHA emergency awards, this would include baseline and endline survey datasets. Partners must provide the AOR with a PDF of the DDL submission confirmation screen as proof of submission of the material to the DDL. Partners should refer to their award documentation for any exemptions, and to the DDL website or ADS 579 for submission requirements.

While BHA recommends submitting non-personally identifiable information (PII), data submitted to DDL can be designated for public publication or not. In order to publish non-PII machine-readable survey data to DDL, the informed consent must indicate that some of the information provided by the respondent will be available on a public website that researchers and others will be able to access without identifying them. See BHA informed consent example in Annex 6.

Resources:

- USAID's ADS 579: USAID Development Data provides policy directives, required procedures, and roles and responsibilities governing the submission of materials to the DDL.

ANNEX 1. SUGGESTED M&E PLAN NARRATIVE OUTLINE

The following is a suggested outline for the M&E Plan to be submitted at Application. Note that BHA accepts that M&E Plans developed for shorter awards (<6 months) may be briefer in detail than those for longer awards. It may also be useful for partners submitting M&E Plans as a deliverable for a cooperative agreement with substantial M&E involvement.

M&E Plan Narrative:

1. **Component 1: Monitoring Approach (required for all awards)**
    a. Specific Data Collection Methods, including for:
        i. Output Monitoring, Outcome Monitoring, Process Monitoring
        ii. Post-distribution Monitoring, if applicable (including sampling)
        iii. Remote Management and Monitoring, if applicable
    b. Context Monitoring
    c. Monitoring Limitations and Mitigating Measures
    d. Data Utilization Plan
    e. AAP Requirement
    f. Data Management and Safeguarding
        i. Data Quality Assurance Procedures
        ii. Data Protection and Security
    g. Staffing and Budget
    h. Abbreviated SOW for Baseline/Endline (see Annex 2 for additional guidance), if applicable
        i. Methods
        ii. Analysis Plan
        iii. Timeframe
        iv. Data sources
        v. Locations
        vi. People responsible
        vii. Limitations and mitigating measures
2. **Component 2: Evaluation Approach (if applicant proposed evaluation)**
    a. Abbreviated SOW for Evaluation (see Annex 4 for additional guidance), including:
        i. Evaluation Purpose
        ii. Evaluation Type
        iii. Evaluation Questions
        iv. Evaluation Methods
        v. Evaluation Timeline
        vi. Evaluation findings dissemination
        vii. Evaluator Profile

ANNEX 2. GUIDANCE FOR ABBREVIATED STATEMENT OF WORK FOR BASELINE/ENDLINE DATA COLLECTION

Applications for emergency activities that are 6 **months or longer** are required to collect baseline and endline data for all indicators. Baseline data are collected in a systematic manner to measure the value of each indicator before the project starts for later comparison. They provide the partner with important information about their affected population that can be used to improve targeting and activity design before implementation begins. The baseline should also describe the prevailing conditions of the beneficiary population and/or situation at the outset of the activity.

This guidance outlines the information to be included in the Abbreviated Statement of Work (SOW) submitted as part of the Application M&E Plan.

Baseline Report: A narrative baseline report and updated indicator tracking table (ITT) with baseline and target values must be submitted to BHA within 90 days of the start of the award.

1. TIMEFRAME

Describe the planned timing for collecting baseline and endline data, including the approximate month. Data collection should take place before implementation has begun in order to get an accurate measure of participants' baseline status, but may coincide with initial implementation where appropriate, such as during beneficiary registration.

If a "rolling" baseline is proposed, please identify when each stage of data collection will occur, and refer to Chapter 5 for additional guidance.

2. LOCATIONS

Present the geographic location for data collection; this should align with intervention areas outlined in the technical narrative of the application.

3. METHODS

Describe the baseline and endline data collection method(s) for all indicators. Methods for baseline and endline should be the same in order to enable comparison. Describe whether the applicant plans to use quantitative, qualitative, or a mixed methods approach. Methods should be appropriate, cost efficient, and in line with humanitarian principles. Data collection methods must adhere to those presented in the PIRS.

In contexts where a partner has back-to-back awards working with the same population, it may be appropriate to use endline data from the previous award as baseline values for some indicators if the

activity targets the same geographic location with similar interventions. Please discuss whether endline data from previous awards will be used as baseline data for the proposed activity.

Many output indicators do not require baseline data collection as their baseline values may be zero. For example, the baseline value for an indicator tracking the number of people trained by the activity is zero.

3.1 Quantitative Methods

Based on the PIRS, identify the indicators for which quantitative baseline and endline data will be collected. Specify whether a survey will be administered directly to beneficiaries (beneficiary-based survey), the general population of the communities being served (population-based survey), or via census. These quantitative methods are described in detail in Chapter 3.

Sampling Plan (if applicant proposes survey): BHA requires probabilistic sampling with PBSs and BBSs. Probability sampling is a selection method whereby every sampling unit within the sample frame has a specific probability of being selected, and that probability can be estimated. For probabilistic sampling, describe the following elements and reference Chapter 3 and the PIRS for more methodological guidance:

a) *Sample frame*: A sample frame is a group of units from which a subset is drawn (e.g., all beneficiaries of an activity or all beneficiaries receiving conditional transfers or all health clinics covered by an intervention or all health clinics in a country). Describe the lists from which primary sampling units (i.e. beneficiaries or households) will ultimately be selected.

b) *Sampling strategy*: The applicant should select from one of the following two strategies: 1) One-stage Simple Random Sample (SRS) (recommended when possible); or, 2) Two-stage Cluster Sampling.

c) *Sample size calculation*: Describe how the applicant will calculate the number of respondents for the survey, and include the confidence level and margin of error. See Section 3.4 for more details on sample size calculation. Discuss whether oversampling will be needed to account for marginalized groups and the level of non-response rate.

3.2 Qualitative Methods

Describe any planned qualitative data collection methods, such as semi-structured in-depth interviews, group discussions, and observation. Qualitative methods may include systematic assessments to shelter, WASH and health facilities, particularly for activities proposing to restore or improve physical infrastructure.

Describe the sampling methods and key attributes to select sample sites and respondents, and estimated number of sample communities, groups, and/or individuals. Describe how the applicant will select sample sites or sample groups. Typically qualitative studies use non-probabilistic sampling methods, such as purposive sampling, but applicants can choose other non-probabilistic sampling methods (e.g., convenience, snowball) depending on the objectives of the study.

4. DATA SOURCES

Specify if primary data will be collected at the population-level of the implementation area or limited to direct beneficiaries and/or other stakeholders (e.g. local authorities and community members). Describe

any secondary data that will be used, such as health facility registries, local market information, local government or administrative datasets.

5. ANALYSIS PLAN

Explain how baseline and endline data will be analyzed and compared. Describe any key analyses that will inform activity targeting and/or implementation. For quantitative surveys, describe how the baseline and endline data will be statistically compared, as appropriate. For some BHA indicators using probabilistic sampling (see Chapter 3), detecting change(s) requires using a statistical package (i.e SPSS, Stata, SAS, CSPro, or other statistical application) and conducting a test of difference. Discussion on the comparison of baseline/endline data should be included in the final performance report, and should be included in the evaluation report if the partner plans to conduct an evaluation, as appropriate.

6. PEOPLE RESPONSIBLE

Identify which position(s) or team(s) will be responsible for gathering the baseline and endline data, and whether data collection will be conducted internally or led by an external consultant. If an external consultant will be hired, please provide a brief summary of the required qualifications.

7. LIMITATIONS AND MITIGATING MEASURES

Describe expected limitations or challenges for data collection. Propose a specific plan or mitigating strategies to overcome each limitation.

8. DATA COLLECTION ETHICS

Describe the applicant's informed consent procedures and the standard operating procedures for ensuring data are secured. This section should also describe how enumerators will be trained in research ethics, including informed consent, and protection of personal information.

ANNEX 3. GUIDANCE FOR ABBREVIATED STATEMENT OF WORK FOR EVALUATIONS

The Evaluation Plan submitted at application must include an abbreviated statement of work (SOW) to allow BHA to assess the technical rigor proposed. The abbreviated SOW should be no more than two pages and address the sections below. The evaluation plan in your application is intended to be a draft outlining your best estimate of what you will evaluate at the time that you are writing the application. Partners must submit a full SOW for BHA review six months prior to the start of the evaluation, which must follow the BHA full evaluation SOW guidance document.

EVALUATION PURPOSE

Briefly describe the purpose of the evaluation and how the results will be used. While an evaluation of the entire activity is acceptable, it is not required; evaluating aspects or components of the activity within a proposed timeframe are also permissible. The following are illustrative examples of evaluation purposes:

a) The effectiveness and relevance of one or more sectoral activities in relation to the activity's goal, purposes, results, and targets.
b) The activity's effects on local markets, and how it affected certain groups of interest (women and men; the youth population; boys and girls, etc.).
c) The effectiveness and relevance of the modality, transfers, and complementary interventions to achieve activity outcomes.
d) Identifying best practices, lessons learned, strengths, and challenges in the activity design, including the LogFrame, and implementation for achieving activity's expected results .

EVALUATION TYPE

BHA supports real-time, formative, and summative performance evaluations at any point during the life of the activity. BHA may also support impact evaluations if the applicant provides a detailed justification of the need for this type of evaluation, which specifically addresses the logistical challenges and ethical considerations that may come with carrying out an impact evaluation in a humanitarian context.

EVALUATION QUESTIONS

Evaluation questions should be relevant to the evaluation purpose and tied to the decisions they are intended to inform. Applicants should limit evaluation questions to five or fewer and questions should be clear, with narrative text or other explanatory information provided to aid understanding. Ensure gender integration into the questions, where appropriate.

Applicants may choose to use relevant OECD DAC [24]evaluation criteria. Some illustrative examples of evaluation questions are presented below, organized by topic:

[24] "OECD Evaluation Criteria", OECD (2020).

a) **Performance**: To what extent have the activity's interventions adhered to planned implementation - schedules, participant targeting, resource transfer composition/quantities, inputs and service delivery, and outputs - and achieved intended goals, purposes and outcomes? Did interventions reach the appropriate target groups and individuals within the target areas? What factors promoted or inhibited adherence to plans and targets?
b) **Effectiveness and efficiency of interventions and their implementation:** To what extent has the intervention appropriately assisted the affected population? How has management adapted the project design or implementation based on monitoring information and feedback from the target population?
c) **Unintended Consequences and Lessons Learned:** What changes—expected and unexpected, positive and negative—did targeted beneficiaries, community members and other stakeholders associate with the activity's interventions? What factors appear to facilitate or inhibit these changes?
d) **Linkages, Layering, and Exit Strategies:** To what extent did the activity take advantage of other USG and non-USG investments in the same space to facilitate linkages with complementary services, layering with earlier investments, and implementing an exit strategy/ies to minimize the dependency on external support? To what extent did the project align and integrate with host government social protection strategy/policy/service delivery?

EVALUATION METHODS & LIMITATIONS

BHA supports evaluations that use qualitative, quantitative, and/or mixed methods. Briefly describe the evaluation methods and ensure that suggested methods are appropriate to the evaluation questions. Please also describe any limitations of the selected methods.

- For quantitative surveys, briefly describe the sampling methodology: will a sample be drawn from the targeted group receiving activity support, or is a population-based survey envisioned (in which any households or individuals living in the target area may be sampled)?
- For qualitative approaches, briefly describe the approach to sampling, e.g., will sample sites or sample groups be selected? BHA encourages the use of a variety of primary data collection methods, including: semi-structured in-depth interviews, focus group discussions, and direct observations (e.g. convenience or snowball sampling).

EVALUATION TIMELINE

The applicant should state the expected period of performance, identifying any specific dates that need to be incorporated in the evaluation plan. Timely scheduling and effective local support contribute greatly to the efficiency of the evaluation team. For evaluations involving complex designs and/or survey research data collection methods, the schedule must allow enough time, for example, to develop sample frames, prepare and pretest survey instruments, train enumerators, and analyze data. Note that all evaluation funding must be obligated during the period of performance of the award.

EVALUATION FINDINGS DISSEMINATION

The applicant should describe the plan for sharing the findings from the evaluation with impacted communities and other stakeholders.

EVALUATOR PROFILE

Briefly describe the intended size of the evaluation team and the specific qualifications that the team members should possess. These skills may include evaluation or methodological expertise, regional or country experience, language skills, management skills, and/or technical subject matter expertise.

BHA requires that the team leader be external to the organization, and encourages evaluation specialists from partner countries to lead or participate in evaluation teams. Where appropriate, BHA staff and/or partners may also participate in the evaluation team. The applicant should describe the intended roles of any participating staff.

ANNEX 4. SUGGESTED FORMAT FOR BASELINE REPORTS

The full report should not exceed 10 pages, excluding the required annexes.

1. **Introduction**
   Describe the award's scope and planned interventions. Please describe the locations and timing of baseline data collection. This should include the objectives of the study and an overview of key findings.

2. **Methodology**
   Provide an overview of the quantitative and qualitative methodology, including a description of sampling (sample frame, sampling strategy, and sample size calculation) as applicable. Please clearly indicate whether any changes in methodology and/or sampling have been made from the approved application Abbreviated Baseline/Endline Statement of Work and provide justification. Describe limitations and mitigating measures taken. If you are using endline data from your previous award as your baseline, please indicate that here. Describe the Applicant's informed consent procedures and the standard operating procedures for ensuring data are secured.

3. **Detailed Findings**
   Describe the prevailing conditions of the beneficiary population(s) including community and/or household characteristics. Describe key findings by sector and sub-sector. Highlight notable differences in baseline values between different segments of the target population by location, age, sex, disability or IDP status, composition of household (i.e. Female & Male Adults; Female Adult No Male Adult; Male Adult No Female Adult; Child No Adult) or other relevant disaggregates.

4. **Programmatic Implications**
   Please describe any adaptations that you will make to your planned activities as a result of the baseline findings, newly identified humanitarian needs or gaps and/or other relevant findings. Highlight and provide justification for any updates to indicator targets from the original application and ensure targets have also been updated in the ITT annex. Note that significant adaptations should be discussed with the AOR.

5. **Conclusion**

6. **Required Annexes**
   Note that annexes do not count towards the 10 page limit.

   a. Indicator Tracking Table, including any proposed updates to indicator targets, as needed
   b. Indicator Estimates Table (only when using representative surveys)
   c. Optional: Survey instruments or data collection tools
   d. Optional: Enumerator Protocol, including Informed Consent

Indicator Estimates Table

The report should include tables with the following information for each applicable indicator:

Indicator	Level of reporting	BL Indicator value	Confidence Interval at 95% level of significance	EL indicator value	Confidence interval at 95% level of significance	# of sampling unit interviewed	in EL, test of difference
FCS (mean or % in Acceptable category)	Overall and disaggregates		± xxx		± xxx		
rCSI (mean)							
HHS (% mod or severe)							

ANNEX 5: SUGGESTED FORMAT FOR BHA EMERGENCY REPORTING

BHA's suggested format for programmatic reporting is adapted from the Grand Bargain 8+3 reporting template and includes USAID, federal and legislatively required components. It includes the eight core questions and one additional question as well as BHA specific instructions.[25] The programmatic performance report should only reflect work done with BHA funding for a specific award number for a specified reporting period.

NARRATIVE COVER PAGE (FOR ALL PROGRAMMATIC PERFORMANCE REPORTS)

I. Reporting Type: *Semi-annual/Annual/Final Performance Report*
II. Reporting FY and Period (dates of reporting period)
III. Partner Name
IV. Award Number
V. Activity Name
VI. Host/Implementation Country
VII. Activity Start Date
VIII. Activity End Date
IX. List of documents uploaded into BHA ART for the reporting period
X. Partner HQ contact person Name, Email, Phone, Office Address
XI. Partner host country contact person Name, Email, Phone, Office Address

---

NARRATIVE COMPONENTS (FOR ALL PROGRAMMATIC PERFORMANCE REPORTS)

1. **Overall Performance:** Write about the activity performance to date. Include information about how successful it is and what results are achieved. Write about the purpose(s) of the activity, and whether or not it is meeting them. Include information about its effect on the different needs of women, men, boys, girls, and vulnerable people. (Suggested length: up to 1 page)

Instructions for partners:

- Write about the activity as a whole but only as an overview. Include information about how progress has been made. Write about the context of the activity, why it was needed, and its original aims.
- It is important to include information about how you found out about the needs of vulnerable people, and how you made sure the activity took their needs into account. Explain how gender considerations were taken into consideration in this activity, and how they were mainstreamed in activity implementation. For example, were men and women involved in the activity design and implementation in a comparable way? Unless the activity was specifically targeted at one group, how did you ensure that men and women benefit from the activity in a comparable way? How did you ensure that the needs and capabilities of persons with disabilities were addressed?

[25] https://www.gppi.net/media/83-Template_final.pdf Note: Grand Bargain 8+3 Reporting Template Core Question 4 Affected Persons is equivalent to the ITT and data entry requirements.

- Please briefly describe your approach to protection. How did you identify risks for affected populations? Which actions did you take to avoid or minimize risks for people?
- This component is different from "Measuring Results". Provide an overview about the activity in general and in relation to its purpose(s). "Measuring Results" component should describe the results in more detail.
- Provide an overall comparison of actual accomplishments on key indicators like beneficiaries reached and value ($USD) value spent by in-kind/cash/voucher modality. Present key results from outcome, output and process indicators that have been collected during this reporting period. For example, this includes data from post-distribution monitoring (PDM), feedback mechanisms and ad-hoc assessments.
- If there is more than one activity sector, describe how they are related with each other and together within the purpose. For example, if there is a food security purpose, describe how the activity sectors are integrated or not. If there is more than one purpose, describe how they are related with each other and together with the goal.
- Describe how the activity has assessed and addressed cross-cutting elements, as applicable, including gender, protection and conflict sensitive needs and issues. If protection issues arise as a direct result of the activity interventions (e.g. increased tension between couples due to women's direct access to cash), the partner must report the unintended effects observed and describe actions taken and planned to mitigate the risk. If language/ethnicity is an applicable cross-cutting element, describe how language/ethnicity is being considered in the activity. For example, have you conducted an assessment or used secondary data to take into account the local languages/ethnicities, literacy rates, preferred local forms of communication, etc.

2. **Changes and Amendments**: Briefly explain any changes to the activity from the original application (whether in the implementation plan, interventions, indicators, or outcomes), and explain why needed to make them, for example because of a change in needs or in the overall situation. (Suggested length: 1/2 page to 1 page)

Instructions for partners:

- Explain any changes or amendments to the original proposal or implementation plan, and the reasons for the changes/amendments. This might include a discussion of how the humanitarian context has changed, changes in the needs of the beneficiaries, or other challenges or problems you had that meant the implementation plan, interventions, indicators, or outcomes had to be changed. If a change was requested and approved by the donor, please mention it.
- Give recommendations for improving the design of the activity or adapting the activity to address these changes, including any changes to activity goals, implementation plan, specific interventions, indicators, or proposed outcomes.
- Highlight major changes in access, security and other relevant elements impacting programming. The description should be specific to changes in the context from the previous performance report, with an outlook at potential changes in the next reporting period. Include reference to any specific issues that have been the subject of informal updates or approval requests to BHA within the last reporting period (e.g., "the cost fluctuations").
- If applicable and commensurate with the importance of market-based programming in your activity, analyze and present results from market assessments and monitoring. Discuss trends and potential impacts on programming, including cost per output, if relevant (price trends should be as compared to normal seasonal fluctuations). You can use data from joint monitoring

systems or your own accurate and up-to-date prices. Also discuss any unexpected market effects and changes made, and if any pre-set thresholds (e.g. price inflation) have been crossed.

3. **Measuring Results**: Describe the progress in achieving the outputs, outcomes and associated targets in the activity application, according to the milestones or indicators that were established. Upload your updated indicator tracking table (ITT); see Upload and Direct Data Entry instructions below. If there are changes to indicator targets and values, data collection methods, and data source, upload the updated ITT and update them in ART(eg. baseline targets and values).

Instructions for partners:

- Write about the outputs, outcomes or results achieved. How much progress has been made towards the targets you identified for each indicator in the original application. Specifically write about whether the targets were met in time, and explain why key targets or milestones were not met, and any differences between the expected results and the actual results. Explain the linkages between outputs and outcomes. Explain the data collection method and verification used.
- Discuss salient activity achievements reflected as planned versus actuals. Discuss and present progress using salient indicators identified in the M&E plan, including narrative description for any results over or under targets.
- Present quantitative and qualitative analyses of outcome, process and output indicators in the narrative, reference any sex-disaggregated results, making sure to address any significant discrepancies in actuals across sexes and by age groups, e.g., if significantly more women were reached than men (and vice versa), or significantly more older women than younger women were reached, and discuss the implication on achieving activity outcomes. Analyses may include data from the ITT, baseline report, the endline report, feedback mechanism, sectoral assessment, markets data and/or PDM report.
- As applicable, BHA requires PDM narrative related to distributions and transfers (i.e., food, non-food items, in-kind, cash, and vouchers), and the role of the goods in achieving the activity purpose(s) and outcomes. The PDM narrative should describe satisfaction with the process of distributions and with the transfers received, as well as beneficiary perspectives on the outcomes of the distributions and transfers.
- Discuss how data will influence technical strategies, interventions, underlying assumptions or the activity's ability to achieve key outputs and outcomes, where applicable.

4. **Participation of and Accountability to Affected Population**: Describe how the activity has been designed to maximize accountability toward the affected population. (Suggested length: 1/2 page)

Instructions for partners:

- How have you given affected populations information about the organization and the activity?
- How have affected populations received information about the partner and the activity? How has this information been well-timed and accessible to everybody? How were people affected by the crisis (including vulnerable and marginalized groups) involved and consulted in the design and implementation of the activity? Which feedback/complaints mechanisms were in place for affected populations to report cases of mismanagement, misconduct and/or sexual exploitation or abuse?

- What did affected persons think about the assistance provided? If possible, quantify beneficiary feedback (for instance "40% of consulted persons find the received support useful"; or 18% of those consulted had complaints").
- How did the partner use beneficiaries' opinions as a guide when making decisions? How was feedback collected, tracked, analyzed and taken into account? Were changes made because of feedback received? If so, how were the changes made? Provide some evidence of collecting and using this feedback (e.g. tools for provision of information, or tracking systems).
- Describe the complaints/feedback mechanisms by providing details about the number of complaints received, proportion of complaints whose issues are resolved, and any salient issues that affect programming and how the information was utilized.

5. Risk Management: Describe how risks to the activity were identified, managed, reduced and mitigated, including any operational, security, financial, personnel management, external or other relevant risks. (Suggested length: 1/2 page)

Instructions for partners:

- Update the risk management analysis and plan included in the application. Were the right risks identified? Were there new risks that the partner did not expect? What were the mitigation measures used to address the identified risks? Did they work?
- Write about external risks from the overall environment, and internal risks, for example, related to financial or personnel management issues. This might include risks of sexual exploitation and abuse of beneficiaries by activity staff, corruption, conflicts of interest, loss of or harm to activity staff, and loss of or harm to activity materials or resources, for example. If the activity takes place in an insecure environment, describe the security risks, including how the security situation evolved over the course of the activity and how this affected activity interventions.
- If applicable, describe all losses regardless of modality must immediately be reported to BHA. When reporting commodity losses include the type, amount and value of commodity including the reason for the loss. For cash and vouchers, losses are defined as any diversion of resource transfers which were intended for participants. When reporting Cash/Voucher losses include value and reason for the loss. The full description of all losses occurring in the reporting period must also be included in the performance report.
- If applicable, describe commodity safety and quality assurance inspection results compared to local country food safety guidance or Codex Alimentarius, as referenced in the Commodity Safety and Quality Assurance section. Results must contain aflatoxin levels and moisture content certification. Commodity safety and quality inspection certificates must be submitted concurrent with performance reports to BHA Partner Portal. Any commodity quality and safety concerns must be immediately reported to the AOR.

6. Coordination: Describe the impact of any coordination efforts, any synergies that developed, and recommendations for improving coordination in the future.

Instructions for partners:

- Describe coordination with the host government, other relevant organizations and the broader humanitarian system, including the cluster system and alignment to HRP/other relevant UN-led appeals/ coordinated responses (where applicable).

- Write about how this has contributed to the activity, for example, any good examples of working together with other interventions, or any other benefits that were the result of coordination. Are there ways that coordination could have been better or could have improved activity outcomes?

7. Challenges and Proposed Solutions: Describe any lessons learned, and how these will be applied to future activities.

Instructions for partners:

- Describe primarily the strongest or weakest parts of the activity, or what parts or strategies made the activity successful or a failure, and explain what you learned from these. Please also reflect on the lessons learned in relation to the activity management, engagement with local partners, your protection interventions, your coordination with affected persons, or to others engaged in the situation.
- Frame responses in terms of what was learned instead of describing what went well or did not go well.

Additional Guidance for Narrative Report

Additional components for SEMI-ANNUAL REPORT

- **Planned Activity.** Provide an overview of key interventions planned for the upcoming reporting period.

Additional components for ANNUAL REPORT

- Write about the activity in the context of the FY timeframe for each narrative component above. Use the unique FY values to support your annual narrative. Unique values avoid double counting across the semi-annual reporting periods. Specific BHA reporting guidance and ART user guide (separate documents) are forthcoming.
- **Planned Activity (if applicable):** For activities that will continue implementation beyond the FY, provide an overview of key interventions planned for the upcoming reporting period.
- **Exit Strategy and Sustainability (if applicable):** Briefly describe the exit strategy and steps to end the activity. Assess the sustainability of its results.
  Instructions for partners:
    - Write about the exit strategy for closing the activity and the expected after-effects of the activity. Focus on the sustainability of the activity, or whether and how results or benefits will continue after it ends.
    - Write about how the activity contributed to the resilience of communities, or how it has supported local partners' capacity. This is particularly important if resilience and support for local partners' capacity were part of the activity application.
    - For some activities, it may also be appropriate to write about ways that parts of it will continue, or will feed into other long-term recovery, rehabilitation or development efforts. For example, did the activity support long-term strategies to reduce humanitarian needs, vulnerability and risks?

Additional components for FINAL PERFORMANCE REPORT

- Write about the activity in the context of the life of award (LOA) timeframe for each narrative component above. Use the unique FY and LOA values to support your final performance narrative. Specific BHA reporting guidance and ART user guide (separate documents) are forthcoming.
- **Activity Overall Performance:** Include information about progress that has been made since the last report. Mention important achievements, problems you have had, or any other information which has affected the activity or its results.
- **Changes and Amendments**: Write about the changes that were made because of the change in circumstances, and how these affected how well you achieved the objectives or milestones set out in the original application.
- **Exit Strategy and Sustainability:** Briefly describe the exit strategy and steps to end the activity. Assess the sustainability of its results. See instructions above under additional AR narrative component description.

Additional BHA Reporting Components

Upload Documents (in BHA ART)

- Narrative Report (Required) - See above suggested format for reporting.
- Indicator Tracking Table (Required) - See ITT for specific instructions.
- Baseline, Evaluation, Assessments, and Research Reports (RiA) - If partner has not yet submitted these reports/documents in BHA ART, do so during as part of the programmatic performance report timeframe. If the report(s) are not finalized and approved by BHA at the time of submission, the partner should note in the narrative that the study/assessment was conducted during the FY, and will upload to BHA ART at a later date.
- Success Stories (recommended) - Success stories are valuable in telling BHA's story. Partners are encouraged to provide this input for public diplomacy and outreach purposes.

Direct Data Entry (in BHA ART)

Partners will directly input semi-annual data in all reporting periods, providing semi-annual and unique FY values (avoiding double counting) for the following:

- Activity level Unique Beneficiaries
- Sector level Unique Beneficiaries
- Activity level Unique Refugee and IDP Beneficiaries
- Emergency indicators

In addition, for the first semi-annual reporting period, partners will provide the following:

- Baseline actual values and LOA target (if have not entered values in BHA ART already)

In addition, for the Final Performance Report, partners will provide the following:

- FY and LOA Activity and Sector Level Unique Beneficiaries
- FY and LOA Level Unique Refugee and IDP Beneficiaries
- FY and LOA Emergency indicators
- Endline/Evaluation values, if applicable

- LOA Actual Data Tables, if applicable
  - LRIP Procurement
  - Modality Actuals

Please refer to BHA reporting guidance and ART user guide (separate document) for detailed information (forthcoming).

ANNEX 6: SUGGESTED INFORMED CONSENT LANGUAGE

BHA recommends using the following informed consent prior to the survey interviews.

Hello. My name is _____. Thank you for the opportunity to speak with you. We are a research team from ___. We are conducting a survey to learn about and try to improve food security, nutrition and wellbeing of your households [Note: Please add other areas that the instrument covers]. Your household has been selected to participate in an interview that includes questions on topics such as your family background, your consumption, food security, and nutrition of women and children. The survey includes questions about the household generally, and questions about individuals within your household, if applicable. The questions about the household and its characteristics will take about 30 minutes to complete. If additional questions are relevant for members of your household, the interview in total will take approximately [xx - adjust based on field testing of questionnaire] hours to complete. Your participation is entirely voluntary. If you agree to participate, you can choose to stop at any time or skip any questions you do not want to answer.

Your privacy is important to us. Private information like your name will not be shared with anyone. Information like information about your consumption may be shared with researchers who will use it to better understand food and nutrition security in your area; these researchers are legally required to protect your information. Some survey responses will also be shared with the public, but no information will be shared that can link you to the study. After entering the questionnaire into a database, we will remove all information such as your name that could link these responses to you before sharing with others for the sake of research.

Do you have any questions about the survey or what I have said? If in the future you have any questions regarding the survey or the interview, or concerns or complaints, we welcome you to contact [your organization], by calling [xxx-xxx-xxxx]. We will leave a copy of this statement and our organization's complete contact information with you so that you may contact us at any time.